

Robert Burduk
Konrad Jackowski
Marek Kurzyński
Michał Woźniak
Andrzej Żołnierek (Eds.)

**Proceedings of the 8th
International Conference
on Computer Recognition
Systems CORES 2013**



Springer

Advances in Intelligent Systems and Computing

Volume 226

Series Editor

J. Kacprzyk, Warsaw, Poland

For further volumes:
<http://www.springer.com/series/11156>

Robert Burduk · Konrad Jackowski
Marek Kurzyński · Michał Woźniak
Andrzej Żołnierek
Editors

Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013

Editors

Robert Burduk
Department of Systems and
Computer Networks
Wroclaw University of Technology
Wroclaw
Poland

Konrad Jackowski
Department of Systems and
Computer Networks
Wroclaw University of Technology
Wroclaw
Poland

Marek Kurzyński
Department of Systems and
Computer Networks
Wroclaw University of Technology
Wroclaw
Poland

Michał Woźniak
Department of Systems and
Computer Networks
Wroclaw University of Technology
Wroclaw
Poland

Andrzej Żołnierek
Department of Systems and
Computer Networks
Wroclaw University of Technology
Wroclaw
Poland

ISSN 2194-5357
ISBN 978-3-319-00968-1
DOI 10.1007/978-3-319-00969-8
Springer Cham Heidelberg New York Dordrecht London

ISSN 2194-5365 (electronic)
ISBN 978-3-319-00969-8 (eBook)

Library of Congress Control Number: 2013939970

© Springer International Publishing Switzerland 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The goal of the CORES series of conferences is the development of theories, algorithms, and applications of pattern recognition methods. These conferences have always served as very useful forum where researchers, practitioners, and students working in different areas of pattern recognition can meet to come together and help each other keeping up with this active field of research. This book is collection of 87 carefully selected works which have been carefully reviewed by the experts from the domain and accepted for presentation during the 8th International Conference on Computer Recognition Systems CORES 2013. We hope that the book can become the valuable source of information on contemporary research trends and most popular areas of application.

The papers are grouped into eight sections on the basis of the main topics they dealt with:

1. Features, learning, and classifiers consists of the works concerning new classification and machine learning methods;
2. Biometrics presents innovative theories, methodologies, and applications in the biometry;
3. Data Stream Classification and Big Data Analytics section concentrates on both data stream classification and massive data analytics issues;
4. Image processing and computer vision is devoted to the problems of image processing and analysis;
5. Medical applications presents chosen applications of intelligent methods into medical decision support software.
6. Miscellaneous applications describes several applications of the computer pattern recognition systems in the real decision problems.
7. Pattern recognition and image processing in robotics which presents pattern recognition and image processing algorithms aimed specifically at applications in robotics
8. Speech and word recognition, which consists of papers focused on speech recognition, automatic text processing and analysis.

Editors would like to express their deep thanks to authors for their valuable submissions and all reviewers for their hard work. Especially we would like to thank Prof. Piotr Porwik and his team from University of Silesia who organized special session devoted to Biometrics, Prof. Jerzy Stefanowski from Poznan University of Technology who helped us to organize special session on Data Stream Classification and Big Data Analytics, Prof. Andrzej Kasiński, Prof. Piotr Skrzypczyński and their team from Poznan University of Technology for their effort to organize special session on Pattern recognition and image processing in robotics. We believe that this book could be a reference tool for scientists who deal with the problems of designing computer pattern recognition systems.

CORES 2013 enjoyed outstanding keynote speeches by distinguished guest speakers:

Prof. Wodzisław Duch - Nicolaus Copernicus University,

Prof. Janusz Kacprzyk - Systems Research Institute, Polish Academy of Sciences,

Prof. Juliusz Lech Kulikowski - M. Nalecz Institute of Biocybernetics and Biomedical Engineering PAS,

Prof. Dariusz Plewczynski - University of Warsaw, Interdisciplinary Centre for Mathematical and Computational Modelling.

Although the last, not least we would like to give special thanks to local organizing team (Robert Burduk, Piotr Cal, Konrad Jackowski, Bartosz Krawczyk, Maciej Krysmann, Bartosz Kurlej, Piotr Sobolewski, Marcin Zmyśloný, Andrzej Żołnierek) who did a great job.

We would like to fully acknowledge support from the Wrocław University of Technology, especially from Dean of Faculty of Electronics, Chairs of Department of Systems and Computer Networks, and The Polish Association for Image Processing which has also supported this event.

Wroclaw,
May 2013

Robert Burduk
Konrad Jackowski
Marek Kurzyński
Michał Woźniak
Andrzej Żołnierek

Contents

Part I: Features, Learning, and Classifiers

Toward Computer-Aided Interpretation of Situations	3
<i>Juliusz L. Kulikowski</i>	
Classification of Multi-dimensional Distributions Using Order Statistics	
Criteria	19
<i>A. Thomas, B. John Oommen</i>	
Knowledge Extraction from Graph-Based Structures in Conceptual	
Design	31
<i>Grazyna Ślusarczyk</i>	
Estimation of the Relations of: Equivalence, Tolerance and Preference	
on the Basis of Pairwise Comparisons	41
<i>Leszek Klukowski</i>	
Generalized Constraint Design of Linear-Phase FIR Digital Filters	53
<i>Norbert Henzel, Jacek M. Leski</i>	
Reduced Kernel Extreme Learning Machine	63
<i>Wanyu Deng, Qinghua Zheng, Kai Zhang</i>	
Using Positional Information in Modeling Inflorescence Discs	71
<i>Małgorzata Prolejko</i>	
A New Method for Random Initialization of the EM Algorithm for	
Multivariate Gaussian Mixture Learning	81
<i>Wojciech Kwedlo</i>	
Heuristic Optimization Algorithms for a Tree-Based Image Dissimilarity	
Measure	91
<i>Bartłomiej Zieliński, Marcin Iwanowski</i>	

Model of Syntactic Recognition of Distorted String Patterns with the Help of GDPPLL(k)-Based Automata	101
<i>Janusz Jurek, Tomasz Peszek</i>	
Decision Rules in Simple Decision Systems over Ontological Graphs	111
<i>Krzysztof Pancerz</i>	
Nonlinear Extension of the IRLS Classifier Using Clustering with Pairs of Prototypes	121
<i>Michał Jezewski, Jacek M. Leski</i>	
Separable Linearization of Learning Sets by Ranked Layer of Radial Binary Classifiers	131
<i>Leon Bobrowski, Małgorzata Topczewska</i>	
Hidden Markov Models For Two-Dimensional Data	141
<i>Janusz Bobulski</i>	
Methods of Learning Classifier Competence Applied to the Dynamic Ensemble Selection	151
<i>Maciej Krysmann, Marek Kurzynski</i>	
The Method of Improving the Structure of the Decision Tree Given by the Experts	161
<i>Robert Burduk</i>	
Part II: Biometrics	
Face Detection and Recognition under Heterogeneous Database Based on Fusion of Catadioptric and PTZ Vision Sensors	171
<i>Aditya Raj, Redouane Khemmar, Jean Yves Eratud, Xavier Savatier</i>	
Eigenfaces, Fisherfaces, Laplacianfaces, Marginfaces – How to Face the Face Verification Task	187
<i>Maciej Smiatacz</i>	
A Content Based Feature Combination Method for Face Recognition ..	197
<i>Madeena Sultana, Marina Gavrilova</i>	
Face Recognition Based on Sequence of Images	207
<i>Jacek Komorowski, Przemysław Rokita</i>	
Statistical Analysis in Signature Recognition System Based on Levenshtein Distance	217
<i>Małgorzata Palys, Rafał Doroz, Piotr Porwik</i>	
A Computational Assessment of a Blood Vessel's Roughness	227
<i>Tomasz Wesołowski, Krzysztof Wrobel</i>	

Performance Benchmarking of Different Binarization Techniques for Fingerprint-Based Biometric Authentication	237
<i>Soharab Hossain Shaikh, Khalid Saeed, Nabendu Chaki</i>	

Biometric Features Selection with k-Nearest Neighbours Technique and Hotelling Adaptation Method	247
<i>Piotr Porwik, Rafal Doroz</i>	

Part III: Data Stream Classification and Big Data Analytics

Predictive Regional Trees to Supplement Geo-Physical Random Fields	259
<i>Annalisa Appice, Sonja Pravilovic, Donato Malerba</i>	

Extending Bagging for Imbalanced Data	269
<i>Jerzy Błaszczyński, Jerzy Stefanowski, Łukasz Idkowiak</i>	

Rule Chains for Visualizing Evolving Fuzzy Rule-Based Systems	279
<i>Sascha Henzgen, Marc Strickert, Eyke Hüllermeier</i>	

Recovery Analysis for Adaptive Learning from Non-stationary Data Streams	289
<i>Ammar Shaker, Eyke Hüllermeier</i>	

Analysis of Roles and Groups in Blogosphere	299
<i>Bogdan Gliwa, Anna Zygmunt, Jarosław Koźlak</i>	

Knowledge Generalization from Long Sequence of Execution Scenarios	309
<i>Radosław Ziembinski (Poznań University of Technology)</i>	

Incremental Learning and Forgetting in One-Class Classifiers for Data Streams	319
<i>Bartosz Krawczyk, Michał Woźniak</i>	

Comparable Study of Statistical Tests for Virtual Concept Drift Detection	329
<i>Piotr Sobolewski, Michał Woźniak</i>	

Part IV: Image Processing and Computer Vision

An Experimental Comparison of Fourier-Based Shape Descriptors in the General Shape Analysis Problem	341
<i>Katarzyna Gościńska, Dariusz Frejlichowski</i>	

Extraction of the Foreground Regions by Means of the Adaptive Background Modelling Based on Various Colour Components for a Visual Surveillance System	351
<i>Dariusz Frejlichowski, Katarzyna Gościńska, Paweł Forczmański, Adam Nowosielski, Radosław Hofman</i>	

Repeatability Measurements for 2D Interest Point Detectors on 3D Models	361
<i>Simon R. Lang, Martin H. Luierssen, David M.W. Powers</i>	
Extended Investigations on Skeleton Graph Matching for Object Recognition	371
<i>Jens Hedrich, Cong Yang, Christian Feinen, Simone Schäfer, Dietrich Paulus, Marcin Grzegorzek</i>	
Low-Level Image Features for Stamps Detection and Classification	383
<i>Paweł Forczmański, Andrzej Markiewicz</i>	
Stochastic Approximation to Reconstruction of Vector-Valued Images	393
<i>Dariusz Borkowski</i>	
Image Segmentation with Use of Cross-Entropy Clustering	403
<i>Marek Śmieja, Jacek Tabor</i>	
Detection of Disk-Like Particles in Electron Microscopy Images	411
<i>P. Spurek, J. Tabor, E. Zajac</i>	
A GPU Accelerated Local Polynomial Approximation Algorithm for Efficient Denoising of MR Images	419
<i>Artur Klepaczko</i>	
Altair: Automatic Image Analyzer to Assess Retinal Vessel Caliber	429
<i>Gabino Verde, Luis García-Ortiz, Sara Rodríguez, José I. Recio-Rodríguez, Juan F. De Paz, Manuel A. Gómez-Marcos, Miguel A. Merchán, Juan M. Corchado</i>	
Real-Time Wrist Localization in Hand Silhouettes	439
<i>Tomasz Grzejszczak, Jakub Nalepa, Michał Kawulok</i>	
An s-layered Grade Decomposition of Images	451
<i>Maria Grzegorek</i>	
System-Level Hardware Implementation of Simplified Low-Level Color Image Descriptor	461
<i>Paweł Forczmański, Piotr Dziurzański</i>	
Reconstruction of Head Surface Model from Single Scan	469
<i>Krzysztof Skabek, Dawid Łapczynski</i>	
The Effectiveness of Matching Methods for Rectified Images	479
<i>Paweł Popielski, Zygmunt Wrobel, Robert Koprowski</i>	
The Print-Scan Problem in Printed Steganography of Face Images	491
<i>Włodzimierz Kasprzak, Maciej Stefańczyk, Jan Popiółkiewicz</i>	

Phototool Geometry Verification	501
<i>Jarosław Zdrojewski, Adam Marchewka</i>	
Structure from Motion in Three – Dimensional Modeling of Human Head	509
<i>Anna Wójcicka, Zygmunt Wróbel</i>	
A Short Overview of Feature Extractors for Knuckle Biometrics	519
<i>Michał Choraś</i>	
Three-Stage Method of Text Region Extraction from Diagram Raster Images	527
<i>Jerzy Sas, Andrzej Zolnierek</i>	
Part V: Medical Applications	
Time Series of Fuzzy Sets in Classification of Electrocardiographic Signals	541
<i>Jacek M. Leski, Norbert Henzel</i>	
Interpolation Procedure in Filtered Backprojection Algorithm for the Limited-Angle Tomography	551
<i>Aleksander Denisiuk</i>	
Classification of Uterine Electrical Activity Patterns for Early Detection of Preterm Birth	559
<i>Janusz Jezewski, Adam Matonia, Robert Czabanski, Krzysztof Horoba, Tomasz Kupka</i>	
Diagnosis of Bipolar Disorder Based on Principal Component Analysis and SVM	569
<i>M. Termenon, Manuel Graña A. Besga, J. Echeveste, J.M. Pérez, A. González-Pinto</i>	
Genetic Algorithms in EEG Feature Selection for the Classification of Movements of the Left and Right Hand	579
<i>Izabela Rejer</i>	
On the Use of Programmed Automata for a Verification of ECG Diagnoses	591
<i>Mariusz Flasiński, Piotr Flasiński, Ewa Konduracka</i>	
Blood Flow Modeling in a Synthetic Cylindrical Vessel for Validating Methods of Vessel Segmentation in MRA Images	601
<i>Grzegorz Dwojakowski, Artur Klepaczko, Andrzej Materka</i>	

Swarm Optimization and Multi-level Thresholding of Cytological Images for Breast Cancer Diagnosis	611
<i>Marek Kowal, Paweł Filipczuk, Andrzej Marciniak, Andrzej Obuchowicz</i>	
Detecting Overlapped Nuclei Regions in the Feulgen-Stained Cytological Smears	621
<i>Bogusław D. Piętka, Annamonika Dulewicz</i>	
Density Invariant Detection of Osteoporosis Using Growing Neural Gas ..	629
<i>Igor T. Podolak, Stanisław K. Jastrzębski</i>	
Cost Sensitive Hierarchical Classifiers for Non-invasive Recognition of Liver Fibrosis Stage	639
<i>Bartosz Krawczyk, Michał Woźniak, Tomasz Orczyk, Piotr Porwik</i>	
Part VI: Miscellaneous Applications	
A Blinking Measurement Method for Driver Drowsiness Detection	651
<i>Belhassen Akrout, Walid Mahdi</i>	
Description of Human Activity Using Behavioral Primitives	661
<i>Piotr Augustyniak</i>	
How to Become Famous? Motives in Scientific Social Networks	671
<i>Adam Matusiak, Mikołaj Morzy</i>	
AdaBoost for Parking Lot Occupation Detection	681
<i>Radovan Fusek, Karel Mozdřen, Milan Šurkala, Eduard Sojka</i>	
Blink Detection Based on the Weighted Gradient Descriptor	691
<i>Krystian Radlak, Bogdan Smolka</i>	
Touchless Input Interface for Disabled	701
<i>Adam Nowosielski, Łukasz Chodyła</i>	
Architecture of the Semantically Enhanced Intellectual Property Protection System	711
<i>Dariusz Ceglarek</i>	
Validation of Clustering Techniques for User Group Modeling	723
<i>Danuta Zakrzewska</i>	
Parking Lot Occupancy Detection Using Computational Fluid Dynamics	733
<i>Tomas Fabian</i>	
Human Fall Detection Using Kinect Sensor	743
<i>Michał Kepski, Bogdan Kwolek</i>	

Evaluation of Various Techniques for SQL Injection Attack Detection	753
<i>Michał Choraś, Rafał Kozik</i>	
Task Allocation in Distributed Mesh-Connected Machine Learning System: Simplified Busy List Algorithm with Q-Learning Based Queuing	763
<i>Agnieszka Majkowska, Dawid Zydek, Leszek Koszałka</i>	
Power Saving Algorithms for Mobile Networks Using Classifiers Ensemble	773
<i>Rafal Lysiak, Marek Kurzynski</i>	
Hardware Implementation of Fourier Transform for Real Time EMG Signals Recognition	783
<i>Andrzej R. Wolczowski, Przemysław M. Szecówka, Jacek Góra, Marek Kurzynski</i>	
Data Preprocessing with GPU for DBSCAN Algorithm	793
<i>Piotr Cal, Michał Woźniak</i>	
Part VII: Pattern Recognition and Image Processing in Robotics	
Structured Light Techniques for 3D Surface Reconstruction in Robotic Tasks	805
<i>M. Rodrigues, M. Kormann, C. Schuhler, P. Tomek</i>	
Multi-modal People Detection from Aerial Video	815
<i>Helen Flynn, Stephen Cameron</i>	
The Classification of the Terrain by a Hexapod Robot	825
<i>Adam Schmidt, Krzysztof Walas</i>	
Robust Registration of Kinect Range Data for Sensor Motion Estimation	835
<i>Michał Nowicki, Piotr Skrzypczyński</i>	
Utilization of Depth and Color Information in Mobile Robotics	845
<i>Maciej Stefańczyk, Konrad Bojar, Włodzimierz Kasprzak</i>	
Part VIII: Speech and Word Recognition	
Texture-Based Text Detection in Digital Images with Wavelet Features and Support Vector Machines	857
<i>Marcin Grzegorzek, Chen Li, Johann Raskatow, Dietrich Paulus, Natalia Vassilieva</i>	

Automatic Disordered Syllables Repetition Recognition in Continuous Speech Using CWT and Correlation	867
<i>Ireneusz Codello, Wiesława Kuniszyk-Józkowiak, Elżbieta Smołka, Adam Kobus</i>	
Evaluation of the Document Classification Approaches	877
<i>Michał Hrala, Pavel Král</i>	
The Prolongation-Type Speech Non-fluency Detection Based on the Linear Prediction Coefficients and the Neural Networks	887
<i>Adam Kobus, Wiesława Kuniszyk-Józkowiak, Elżbieta Smołka, Ireneusz Codello, Waldemar Suszyński</i>	
Author Index	899

Part I

Features, Learning, and Classifiers

Toward Computer-Aided Interpretation of Situations

Juliusz L. Kulikowski

Abstract. The problem of interpretation of situations as a widely extended and important component of living beings' behavior in real world is considered. A scheme of interpretation of situations in natural living beings is presented and a general scheme of inspired by the nature artificial situations interpreting system is proposed. Basic constraints imposed on computer-based situations interpreting systems are described. It is shown that the computer-based situations interpreting systems are an extension of pattern recognition systems and the differences between them are characterized. The role of domain ontologies and of ontological models in computer-based situations interpreting systems design is shown and it is illustrated by examples.

1 Introduction

70 years since the publication by W.S. Mc Culloch and W.H. Pitts of their concept of a *mathematical model of neuron* in 2013 will be passed. Their significant publication initiated an increasing interest of scientists and of engineers to construct mechanisms simulating functional abilities of a natural nervous system. A next important step in this direction was made by F. Rosenblatt in the late 50ths of the past century by publication of his concept of *perceptron* - a first *artificial neural network* recognizing simple graphical patterns. This was a starting point of *pattern recognition* as a new scientific discipline which for the following years resulted in numerous new concepts, original computer programs and many thousands of considerable papers. Since that time, pattern recognition evolves in two competing streams of works: the first one based on *artificial neural networks* (ANN)

Juliusz L. Kulikowski

Nalecz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences,
4 ks. Trojdena Str., 02-109 Warsaw, Poland

e-mail: juliusz.kulikowski@ibib.waw.pl

and the second one using various algorithmic (statistical, functional, algebraic, formal linguistic, etc.) approaches to classification and recognition of patterns. In fact, the first approach, which originally tried to construct pattern recognition systems in hardware, finally realize them in the form of computer programs simulating artificial neural networks. Therefore, both approaches in some sense are "algorithmic", the main difference among them consisting in the sources inspiring with the concepts of algorithms: modeling of neural networks or modeling of formally defined classes of objects' similarity. Against original expectations, artificial neural networks rather few contributed to our understanding of the natural thinking mechanisms. This is rather the progress of neuropsychology and neurophysiology which for the last decades provided new concepts which may be useful in improving and extending the intelligent computer-aided data processing methods and algorithms. In particular, a lot of highly effective pattern recognition methods and computer programs in application to detection and recognition of simple printed or hand-written characters, voice signals, defects of machine details, diseases symptoms, biological cells, finger-print minutiae, etc. became available and are presently widely used. On the other hand, the problems of computer-aided image understanding on advanced semantic level, interpretation of observable situations connected with their pragmatic evaluation, early natural disasters prognoses, effective recognition of admissible ways to reach a goal in unclear circumstances, etc. still remain to be solved, despite the fact that our natural mind less or more effectively solves them everyday. Let us remark that the above-mentioned problems go beyond the limits of the conventional pattern recognition ones, however, they still concern recognition in a more general sense: assigning of some widely defined "objects" to fuzzy similarity classes in connection with their semantic and/or pragmatic assessment. Moreover, till now, no effective results of image, scene or situation analysis by artificial neural networks in the above-mentioned, wide sense have been achieved. This is clear because architecture of no artificial neural network (see, e.g. [1], [2], [3]), sophisticated as might it be, reminds the real functional structure of a natural brain (see, for a comparison [4], [5]). This does not mean that the state of ANN will not be evolving; nevertheless, it seems that in the nearest future more effective methods of composite objects recognition and interpretation will be provided by the approaches based on formal models. This expectation follows from the fact that a lot of computer-aided decision-making systems already exist; however, the role of computers in this type of systems is usually reduced to acquisition, primary processing, partial analysis, presentation and/or storage of data for final decision making by a human user. The user is faced with a deontological problem [6]:

If it is known that a situation A arose then undertake an action B in order to reach a more desirable situation C.

In this context, our problem consists in designing algorithms stating that a situation A really arose; this is a general objective of the *situation interpretation* theory and practice. At a first glance, it might seem that *pattern recognition* is a discipline solving the above-mentioned problem satisfactorily. E.g., if in a given histological

specimen cancer cells have been recognized then usually it follows from this that some therapeutic actions should be undertaken in order to cure the patient. However, the cancer cells are in this case no more but a component of a more general situation: they have been detected in an inner organ of a patient of a given gender, age, etc. whose less or more extended and accurate case record is known. All this information constitutes a situation a medical doctor has to deal with. Therefore, *pattern recognition* can be considered as a particular case of *situation interpretation*. Nevertheless, the notions of *situation* and *interpretation* remain to be more strictly defined. In this paper an attempt to the formulation of *situation interpretation* (SI) backgrounds as well as some suggestions concerning the computer-aided SI systems design are presented. We try to show that basic notions of SI can be founded on well known mathematical concepts on which ontological models are based while practical solution of SI can be inspired by the natural cognitive processes created by the evolutionary processes.

2 Basic Notions

2.1 Natural Interpretation of Situations

There are two basic approaches to definition of the notion of *situation*. In a widely used sense **situation is a passing state of a selected fragment of reality**. Reality can thus by a subject be considered as a stream of surrounding her/him or it, arising, evolving, affecting some other ones or being by some other ones affected and finally disappearing situations. Any living being deals with various situations as an observer, their active subject, object, commentator, etc. However, situations are not the components the reality consists of; they rather are the components of perception of the reality by living beings. The same objects or fragments of real processes can by different subjects be considered as elements of different situations. This follows from a general, in fact - subjective way a living being learns to perceive and to discriminate the situations. E.g., the same state considered by a subject in the context of some preceding and some following states can be assessed as "desired" or "undesired" and as such it will be differently interpreted. The way of situations recognition by simple living beings is illustrated in Fig. 1.

They reach their ability to perceive situations by perception of various temporary signals coming from the real world, primary assessment of their value as "more" or "less" agreeable, remembering, collecting according to some similarity features, discriminating their components, associating according to their time- and/or space-co-occurrence and assigning to them specific reactions in the form of signals or behaviors expressing subjects' emotions caused by the situations. According to our present knowledge about natural brain functions, most of the above-mentioned operations using memory takes place in an anatomic part of the brain, belonging to its temporal lobe, called a hippocampus. Emotional values to the impressions are assigned in adjacent to it amygdale body. The emotionally marked reactions to many

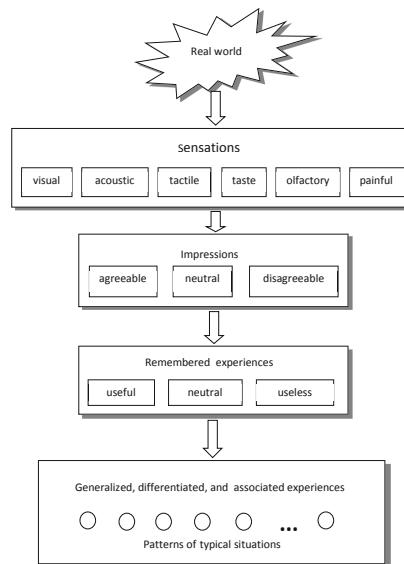


Fig. 1 A natural way to recognize situations in simple living beings

times in similar situations repeated impressions become established as germs of a code used to communicate about the situations by the members of a social group of living subjects. Therefore, recognition of situations even in simple living beings is closely connected with assessment of their importance for the recognizing subjects. Otherwise speaking, it contains some aspects of their "interpretation" even if not expressed in terms of any advanced language.

In more developed living beings the natural way of situations interpretation is in general similar, however, some differences should also be remarked:

- Instead of simple, sensations some formerly established sub-situations can be used as a basis to construct the concepts of higher-level situations;
- The scale of situations' practical importance assessment becomes more sophisticated, also in a multi-aspect sense;
- Linguistic terms are assigned to typical situations;
- Elements of classification of the situations may be used;
- Situations are assessed also from a point of view of their relations to other: past, existing, expected or possible situations.

In a large sense, **interpretation of situation** means a natural cognitive process consisting of:

- acquisition of observations;
- primary situation recognition;

- situation's position localization in a more general system of concepts;
- establishing of relations between the given and other (past, present, future possible, etc.) situations.

As a result, this leads to an ability to avoid emergency, to find food, to contact other social partners or to make another decision of vital importance.

A scheme of advanced natural situations interpretation is shown in Fig. 2. In fact, the scheme is rather hypothetical, because till now it is rather few known about the natural processes and mechanisms of situations interpretation in primates. However, a progress in modern brain imaging modalities (e.g. functional magnetic resonance, f-MRI [6], diffusion tensor imaging, DTI [7]) opens new possibilities in this research area.

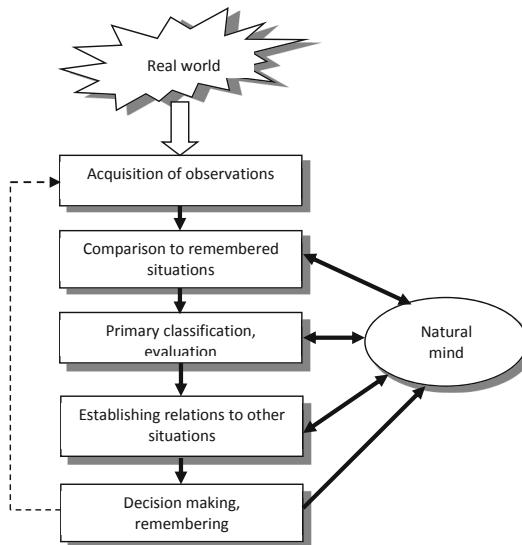


Fig. 2 Advanced natural interpretation of situations

2.2 Formal Model of Situation Interpretation

A general architecture of a computer-based situations interpretation (SI) systems can be inspired by the scheme shown in Fig. 2, excepting that the "natural mind" should be replaced with computer-based knowledge base and logical engines. However, a strong simulation of the natural way of learning the recognition and interpretation of situations for certain reasons seems to be unrealistic. First, because of a long, hundreds thousands years counting duration-time of the natural evolutionary processes. Second, because our knowledge about the natural mechanisms of everyday experiences acquisition, association, generalization, evaluation, storage and reminding is

still now incomplete. Therefore, the computer-based SI system cannot be modeled on the natural one; they rather should be based on an abstract idea of ontology which in the last years in information processing systems design became popular [8], [9].

A general ontology is usually defined as a part of metaphysics concerning the nature and theory of existence. In computer applications an idea of domain ontologies defined as "a common understanding of some domain" [8] or as "an abstract view of the world we are modeling, describing the concepts and their relationships" [9] became more useful. Any domain ontology defines a hierarchy of its concepts representing objects, attributes, states, actions etc. used to the description of the given part of reality. In [10] a domain ontology decomposition into an ensemble of ontological models was proposed:

$$O = \{OM_1, OM_2, \dots, OM_I; Q, A\} \quad (1)$$

where any ontological model is defined as a quadruple:

$$OM_i = \{C_i, R_i, T_i, A_i\}, \quad i = 1, 2, \dots, I. \quad (2)$$

Above, C_i denotes a non-empty family of concepts (objects, attributes, actions, etc.) the OM is based on; R_i is a family of relations described on selected subsets of C_i . Among the relations in R_i a multi-aspect taxonomy Ξ_i of the concepts is distinguished, a concept $T_i \in C_i$ being its top-level element, the highest concept of the OM_i .

Basic formal properties of the ontology as well as of its ontological models follow from the corresponding sets of axioms A and A_i -s. In particular, the axioms of the algebra of sets, of relations and of their extensions theory, of fuzzy sets, of probability theory, etc. if used as bases of the given OM 's definitions should be considered as components of their formal description.

A general structure of the ontology is described by a super-relation Q between the OM s. In particular, clustering of OM s into categories corresponding to various aspects of the domain reality suits to putting an ontology in a clearer order. Despite the fact that in practice, the axioms in the description of the ontological models are very often neglected, they should be strongly respected if an ontology is to be correctly designed.

Sequences of elements of C_i satisfying a given relation $\rho \in R_i$ are called *syndromes* of ρ . Similarity relations describe particular types of concepts called *patterns*. OM s consisting of similarity relations and of some based on them higher-order relations describe a particular type of syndromes called *scenes*. Otherwise speaking, a scene is a composition of objects representing some patterns. As such, it can be described in terms of concepts of a domain ontology.

Sets of syndromes commonly satisfying the relations of a given OM , more generally, will be called situations of the model. Observed scenes are thus

particular cases of observed situations; recognition of scenes or of patterns is a particular case of **situation recognition**. However, **interpretation of situation** consists of:

- situation recognition,
- its pragmatic aspects assessment,
- its logical extension.

The last term denotes detection within the given domain ontology other *OMs* whose syndromes may be logically consistent with the analyzed situation. Logical extension makes thus possible answering questions not only directly concerning the analyzed situation but also some questions concerning its possible connections with other situations.

Not all domain ontologies and ontological models can be used to situations interpretation. For this purpose, they should contain a *mechanism V* of the situations' *pragmatic aspects assessment*. It should provide a possibility to compare any syndromes from the point of view of their significance, utility, emergency level, etc. for the observer. In the simplest case it can take the form of a binary scale (e.g., "important", "not important"). Therefore, an *OM* admitting pragmatic evaluation of situations takes the form of a quintuple:

$$EOM_j = [C_j, R_j, T_j, V_j, A_j], \quad j \in [1, 2, \dots, I], \quad (3)$$

(C_j, R_j, T_j, A_j denoting the same objects as in (2)). This type of *OMs* can be called *evaluated ontological models*.

The above-presented elements of formalism lead to a hierarchy of real world representing models suitable to be used in a computer-based *SI* system, consisting of:

- *a general ontology*;
- *domain ontologies*;
- *a hierarchy of ontological models* describing the inner structure of the domain ontology as a composition of ontological models;
- *ontological models (OM, EOM)* describing selected aspects or states that in a given domain may occur and can be observed;
- *relations, super-relations or hyper-relations* of any type specifying the ontological models;
- *situations* described by the ontological models;
- *formal patterns* specifying the *similarity classes, structures* satisfying *ordering relations* and *syndromes* satisfying other types of relations, super-relations and/or hyper-relations;
- *sets of objects, values of attributes, alphabets* etc. on which the patterns, formal structures, symbolic expressions, etc. are defined.

The multi-level structure of models and their components is shown in Fig. 3.

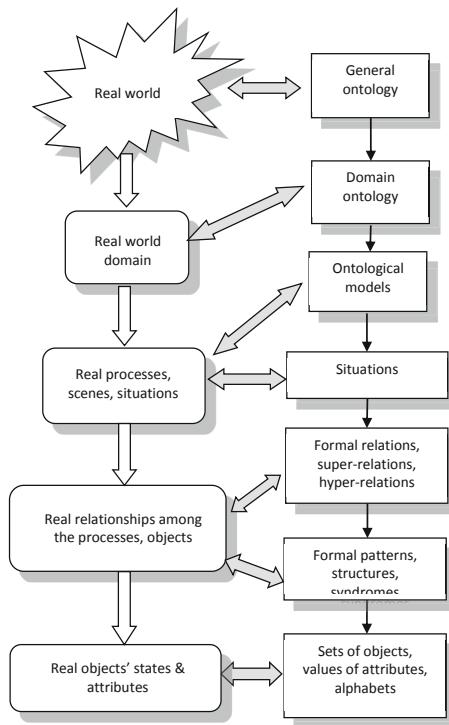


Fig. 3 Relationships between real world and its formal models

3 Evaluated Ontological Models

3.1 General Remarks and Constraints

From a practical point of view, *EOMs* are a form of acceptable by computers description of the situations that in a given application domain may arise. A repertoire of formal tools suitable to construction of *EOMs* is thus limited to those expressible by computer languages or data structures. Some sophisticated mathematical notions like: a family of all subsets of a continuous set or a set of all transcendental numbers between 0 and 1, etc. cannot in computer systems exactly be represented. Therefore, *OMs* or *EOMs* on this type of incalculable notions cannot be based. As a consequence, not all real situations can by ontological models be described.

More generally, it seems impossible to recognize and to interpret situations in the case of:

- real situations not expressible in correctly defined and calculable formal terms;
- inadequacy of the concepts of a domain ontology to real input (observation) data;

- lack of effective algorithms of processing on formal terms the ontological models are based on;
- lack of adequately chosen methods of situations similarity and pragmatic aspects assessment.

As examples of real situations that for the above-listed reasons (till now) cannot by a computer-based *SI* system be analyzed the following situations can be mentioned:

- connected with subjective psychical impressions or consciously non-controlled emotions (e.g. impression of fear evoked by a thriller emitted in tv, people's excitement caused by a street accident, etc.);
- described by logically incompact datasets (e.g. a total welfare level which in different countries is evaluated by different methods);
- concerning real life domains not described by strongly defined or measurable terms (e.g. ambience at a board meeting);

etc. Despite the above-mentioned limitations, an area of real *SI* problems possible to be solved by using available mathematical tools remains still large.

Design of an *EOM* for a given *SI* purpose can be facilitated by a preliminary answering the following questions:

- Concerning the domain:
 - What is the name of the area of interest;
 - Does it exist a domain ontology for it? (if not, start to construct it);
- Concerning a general structure of the domain:
 - What is the name of the specific sub-area of interest;
 - Does it exist a sub-ontology for it?
- Concerning the ontological model:
 - What is the name of the situation of interest within the ontology or sub-ontology;
 - Does it exist an ontological model for it?
- Concerning an "anatomy" of the situational ontological model:
 - What are the active and/or passive objects participating in the situation?
 - What are the attributes of the objects?
 - What are the types of substantial relations among the objects?
 - Are there any typical examples of instances of the relations?
 - What are the criteria for other objects' approval as instances of the relations?
 - What features of the relations' instances should be taken into account for their pragmatic value assessment?
 - What are the relations whose instances' time-evolution for pragmatic value assessment should be taken into consideration?
- Concerning the super-relations within the domain:

- Are there any other types of situations that in some way may be associated to the given one (e.g. coexisting in time, influencing one the other one, etc.)? Name them.
- Are the other types of situations described by ontological models?
- Is it formally possible to define a super-relation on the basis of the associated relations?
- What pragmatic values can be assigned to the instances of the super-relation as a function of pragmatic values of its components?

3.2 Categorization

Besides the above-mentioned super-relations described on some combinations of relations, the last can be categorized according to some criteria following from a general domain structure. E.g., within a domain *Health service* the following categories of *OMs* can be specified:

- topographical: *Warsaw City, South region, Country*, etc.;
- functional: *Diagnosis, Treatment, Education, Prevention*, etc.;
- medical: *Internal diseases, Psychiatry, Oncology*, etc.;
- organizational: *Hospitals, Outpatients clinics, Ambulance stations*, etc.

Evidently, the taxonomy of *OMs* may be horizontally or vertically-down extended, e.g.:

Warsaw Center Home visit rounds, Influenza, Private outpatient clinics.

This may associate by a super-relation an *OM* satisfying the above-given criteria: modeling all *Private outpatient clinics* in *Warsaw City* offering *home visits* in the case of *influenza*.

Categorization of *OMs* makes also possible recognition of "higher level" situations like: "*Abnormal number of patients registered on Monday in all public health service centers of Warsaw City with the symptoms of influenza*" and assigning to them pragmatic values like: *important* or *alarming*. This is possible if in the *OM* among the relations describing all types of *Health service centers* there are the ones representing an ontological concept *State of services* described by the following lower-level concepts (objects, attributes):

(Health service center), Home visit #, Date, Case (medical diagnosis)

On the basis of all relations of this type in all *OMs* describing *Health service centers* in *Warsaw City* it can be constructed a super-relation defined as an extended algebraic sum [12] of sub-relations obtained by fixing the data:

Date := Monday, Case := influenza.

Such a super-relation will be satisfied by all syndromes of the form:

**, *, Monday, influenza*

where the unknown values of *Health service center* and *Home visit* \ddagger have been denoted by *).

3.3 Pragmatic Value Assessment

From a formal point of view, a mechanism of situations' pragmatic value evaluation is a sort of a *weak semi-ordering relation* V imposed on the situations of a given *OM*. Before it to be defined, let us remind the notion of a *strong similarity* as a reflexive, symmetrical and transitive relation \approx [13]. Then, if it is given a set of situations admitted by the *OM* and ω' , ω'' , ω''' are some particular situations then V should satisfy the following conditions:

- V is satisfied by any pair $[\omega', \omega']$ (reflexivity);
- If V is satisfied by an $[\omega', \omega'']$ then it is also satisfied by $[\omega'', \omega']$ if and only if $\omega' \approx \omega''$ holds (weak asymmetry);
- If V is satisfied by an $[\omega', \omega'']$ and by $[\omega'', \omega''']$ then it is also satisfied by $[\omega', \omega''']$ (transitivity).

We call *equivalence classes* the sets of objects mutually equivalent in the sense of the (mentioned in the definition) strong similarity (\approx) relation.

The mechanisms of pragmatic value evaluation into two groups can be divided:

- Single-aspect evaluation (*SAE*) mechanisms,
- Multi-aspect evaluation (*MAE*) mechanisms.

SAE mechanisms can directly be based on the above-given definition of weak semi-ordering relations. They may suit to evaluate such situations' properties as *importance*, *usefulness*, *level of interest*, *emergency level*, etc. However, they may take the forms of:

- a. strong (linear) ordering;
- b. semi-ordering;
- c. partial semi-ordering

Strong linear ordering means that for any different situations ω' , ω'' the relation V is satisfied either by $[\omega', \omega'']$ or by $[\omega'', \omega']$; however, the equivalence classes may contain no more but one element. In "standard" (denoted b)) and partial (denoted c))semi-ordering multi-element equivalence classes are admissible. However, partial semi-ordering admits also that for certain pairs of situations the relation V is neither

by $[\omega', \omega'']$ nor by $[\omega'', \omega']$ satisfied. This is illustrated in Fig.4 a, b, c, where circles represent situations, arrows - increasing pragmatic values and multi-element equivalence classes are by dotted contours denoted.

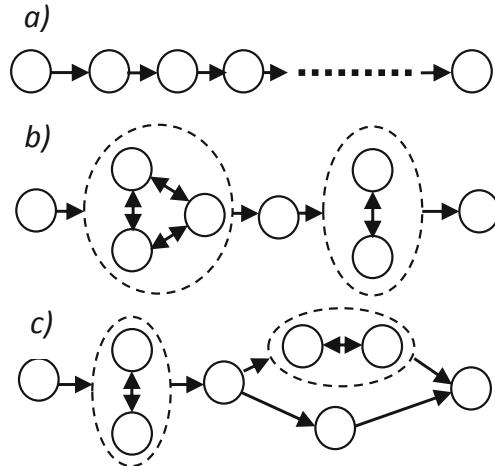


Fig. 4 Types of pragmatic values ordering: a) strong (linear), b) semi- ordering, c) partial semi-ordering

In the cases a) and b) numerical or nominal scales of pragmatic values can be used while in the case c) pragmatic values by their relative comparison only can be assessed.

MAE mechanisms on the basis of particular *SAE* mechanisms can be defined. One of the simplest ways to reach this is introducing a higher-level lexicographic order between the *SAEs*. For this purpose, let the set of particular *SAEs* be linearly ordered as $[SAE^{(1)}, SAE^{(2)}, \dots, SAE^{(k)}]$. Then, for a given pair of situations $\{\omega', \omega''\}$ to be compared their order is first established in $SAE^{(1)}$. If they belong to different equivalence classes then the corresponding order is to them assigned and the procedure is finished. Otherwise, if they belong to the same equivalence class, their order is established according to the $SAE^{(2)}$ rule, etc., up to reaching $SAE^{(k)}$.

4 Interpretation of Situations

4.1 Recognition of Situation

Let u denote a sequence (ordered set) of results of observations in a certain application area. The elements of u may be of **various physical and/or formal nature**.

Recognition of a represented by \mathbf{u} situation consists in **finding in a corresponding domain ontology a model** OM_i whose relations are by \mathbf{u} (or by its subsets) fulfilled. Apparently, it reminds a typical pattern recognition problem in which the role of OMs is played by some predefined patterns [14]. However, the differences between situations and pattern recognition are substantial:

- the universe U of all possible observations usually does not constitute any homogenous, well-defined mathematical space;
- the observations $\mathbf{u}, \mathbf{u} \in U$, should fulfill several relations of various nature;
- relations can be characterized analytically, by examples or by logical tests [15];
- the number of relations in different OMs may be different;
- situation recognition is a multi-step decision process controlled by pragmatic assessment of the results reached at the preceding steps.

Recognition of a situation needs a preliminary analysis of U from a point of view of its **formal and semantic consistency** with the concepts defined in the domain ontology. The mentioned in Sec. 3 basic constraints imposed on OMs should also be obeyed. Moreover, in the case of OMs based on random or fuzzy relations the fact that they are by \mathbf{u} fulfilled can be expressed not only in a binary ("yes", "no") logical scale but sometimes also in a continuous numerical scale of probability or of a membership measure. It thus arises a problem, **how to define a total measure of the fact that a given OM is by \mathbf{u} fulfilled as a function of various particular measures of the corresponding OMs relations being by \mathbf{u} fulfilled**.

The problem can be solved by imposing some constraints on the form of a fulfilling measure $\varphi_\rho(\mathbf{u})$ of a relation being fulfilled by the syndrome \mathbf{u} . It will be assumed that in general $\varphi_\rho(\mathbf{u})$ is a function satisfying the conditions:

1. $0 \leq \varphi_\rho(\mathbf{u}) \leq 1$
2. $\varphi_\rho(\mathbf{u}) = 0$ if \mathbf{u} does not satisfy ρ (also, if ρ is an empty relation);
3. for any \mathbf{u} if $\rho' \subseteq \rho''$ then $\varphi_{\rho'}(\mathbf{u}) \leq \varphi_{\rho''}(\mathbf{u})$

The following types of measures can be taken into consideration:

1. based on discrete scales 0 - does not satisfy, 1, 2, ..., k - certainly satisfies; $k = 1$ (binary), $k = 2$ (ternary), etc.;
2. based on a continuous scale $[0, 1]$.

Then, if $\varphi_0(\mathbf{u})$ denotes a total measure and $\varphi_n(\mathbf{u}), n = 1, 2, \dots, N$, the measures of particular relations being satisfied by \mathbf{u} , it can be assumed that:

$$\varphi_0(\mathbf{u}) = \varphi_1(\mathbf{u}) \cdot \varphi_2(\mathbf{u}) \cdot \dots \cdot \varphi_N(\mathbf{u}). \quad (4)$$

Example 1 It is given an observation:

*Date / Hour / Name / Address / Age / Sex / Diagnosis / State
21.03 / 18.00 / N.N. / aaa / 67 / male / apoplexy / suspected*

A doctor called to the urgent case needs a remote medical consultation of this case with a neurological specialist. For this purpose, he should introduce his observation data to a *Medical service OM* containing the relations:

*Medical center, Center's address, Center's e-mail;
 Center's departments, Center's departments' phones;
 Center's department's specialists;
 Center's department's duty schedule.*

This makes possible to establish a relation:

Date / Hour / Diagnosis / State / Medical center / Center neurological department / Center neurological department's phone / Neurological specialist on duty.

Finally, this leads to recognition of a situation like:

21.03 / 18.00 / N.N. / aaa / 67 / male / apoplexy / suspected / Medical center XX / Neurological department / Phone number zzz / Neurological specialist dr YY/.

4.2 Interpretation of Situation

The above-presented example might suggest that *SI* is a simple one-step decision process. In fact, it is an iterative process consisting of several steps, as it has been mentioned in Sec. 2.2. This can be illustrated by a next example.

Example 2

Let us take into consideration a computer-aided street-traffic monitoring system. Its tv camera has caught a series of shot pictures illustrated schematically in Fig.5.

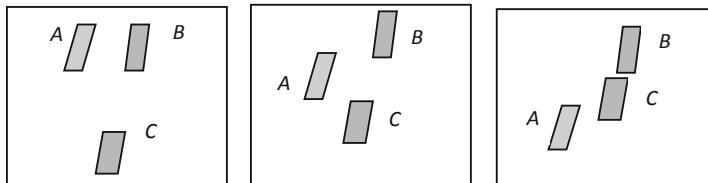


Fig. 5 Three consecutive shot pictures of a street-traffic film

Let us assume that three objects (denoted *A*, *B*, *C*) have been detected in the pictures. A preliminary comparison of the pictures leads to finding that:

A is approaching, B is steady, C is going away

The aim of the *SI* system is to detect and to record traffic accidents. The interpretation process may run as shown in Fig.6.

In the above-described process a following qualitative linear scale of values has been used:

Not important → Possibly important → Suspected → Important → Highly important

Observations	Recognition	Evaluation
Objects A, B, C	A scene A-B-C	Possibly important
Objects are moving	A moving scene A-B-C	Important
A beside B, A closer to C B closer to C	No conflict A-B Conflict A-C, Conflict B-C	Not important A-B Suspected A-C Suspected B-C
Shapes A, B, C	A – small vehicle, B – small vehicle C- big vehicle	Important A-C Important B-C
A beside C, B very close to C	No conflict A-C Conflict B-C	Not important A-C Important B-C
B touches C	Accident B-C	Highly important B-C

Fig. 6 Street traffic monitoring process

The *SI* process contains several pattern recognition acts: recognition of a *scene* (general), *moving scene*, *conflict*, *vehicle*, *accident*, etc. The mentioned here types of recognized objects as some ontological concepts are described by corresponding *OMs*. For example, the concept of *accident* can be defined by a hyper-relation H_Q described on a set Q consisting of at least two *objects* localized in a metric space and constructed as follows:

1. It is taken into account a Cartesian product $Q \times Q$;
2. It is defined a relation $r \subset Q \times Q$ as a set of all pairs of different elements of Q ;
3. It is defined a table D_t of *metric distances* between the pairs of elements of r ;
4. There are taken into consideration the tables D_{t1}, D_{t2}, D_{t3} for three consecutive time-instants $t_1 < t_2 < t_3$;
5. H_Q is fulfilled if for at least one pair of objects $q', q'' \in Q$ their metric distances satisfy the relation:

$$0 < dt_1(q', q'') > dt_2(q', q'') > dt_3(q', q'') = 0. \quad (5)$$

5 Conclusion

Interpretation of situations is a substantial component of living beings' behavior in real world. It consists of recognition of arising situations, their pragmatic assessment and establishing their relations to other situations. Similar type of decision making processes can be modeled in computer systems. Computer-based *SI* systems cannot

directly simulate the natural processes but they can be by them inspired. For this purpose domain ontologies and ontological models can be used. In the paper a general approach to the *SI* systems design has been proposed. There also have been indicated some basic constraints imposed on the artificial *SI* systems. In general, they are connected with incalculability of some formal (mathematical) concepts. Nevertheless, the ideas of extended algebra of relations and of super- and hyper-relations theory, including their fuzzy extensions, as well as suitable domain ontologies and ontological models are powerful tools for realization of effective computer-based *SI* systems for various important applications.

References

1. Hertz, J., Krogh, A., Palmer, R.G.: Introduction to the theory of neural computation. Addison-Wesley Publishing Co., Redwood City (1991)
2. Levy, B.C., Adams, M.B.: Global optimization with stochastic neural networks. In: IEEE Conf. on Neural Networks, San Diego, USA, pp. 681–689 (1987)
3. Hopfield, J.J.: Artificial neural networks. IEEE Circuits and Devices Magazine, 3–10 (1988)
4. Rilling, J.K., Insel, T.R.: The primate neocortex in comparative perspective using magnetic resonance imaging. Journal of Human Evolution 37, 191–223 (1999)
5. Allen, J.S.: The lives of the brain. The Harvard College (2009)
6. Rademacher, J., Galaburda, A.M., Kennedy, D.N., et al.: Human cerebral cortex: Localization, parcellation and morphometry with magnetic resonance imaging. Journal of Cognitive Neuroscience 4, 352–374 (1992)
7. Ramnani, N., Behrens, T.E.J., Johansen-Berg, H., et al.: The evolution of prefrontal inputs to the corticopontine system: Diffusion imaging evidence from macaque monkeys and humans. Cerebral Cortex 16, 811–818 (2006)
8. Fernandez-Lopez, M., Gomez-Perez, A.: Overview and analysis of methodologies for building ontologies. The Knowledge Eng. Rev. 17, 129–156 (2002)
9. Gruber, T.R.: A translation approach to portable ontologies. Knowledge Acquisition 5, 199–220 (1993)
10. Kulikowski, J.L.: The Role of ontological models in pattern recognition. In: Kurzynski, M., et al. (eds.) Computer Recognition Systems. Advances in Soft Computing, pp. 43–52. Springer, Berlin (2005)
11. Kulikowski, J.L.: Relational approach to structural analysis of images. Machine Graphics and Vision 1, 299–309 (1992)
12. Kulikowski, J.L.: Pattern recognition based on ambiguous indications of experts. In: Kurzynski, M. (ed.) Komputerowe Systemy Rozpoznawania, KOSYR 2001, Wroclaw, pp. 15–22 (2001)
13. Owsinski, J.W., Brüggemann, R. (eds.): Multicriteria ordering and ranking: partial orders, ambiguities and applied issues. System Research Institute, Warsaw (2008)
14. Duda, R., Hart, O., Stork, D.: Pattern classification. John Wiley, New York (2000)
15. Przytulska, M., Kulikowski, J.L., et al.: Computer methods of radiological images analysis for pathomorphological lesions assessment in selected inner organs. Report on the project N N518 4211 33 IBBE PAS Warsaw (2010)

Classification of Multi-dimensional Distributions Using Order Statistics Criteria

A. Thomas and B. John Oommen*

Abstract. This paper proposes a novel classification paradigm in which the properties of the *Order Statistics* (OS) have been used to perform an optimal/near-optimal solution for multi-dimensional problems. In our initial works in [5] and [6], we proposed the foundational theory of CMOS, Classification by the Moments of Order Statistics, for some uni-dimensional symmetric and asymmetric distributions of the exponential family. In this paper, we generalize those results for various multi-dimensional distributions. The strategy is analogous to a Naïve-Bayes' approach, although it, really, is of an *anti-Naïve-Bayes'* paradigm. We provide here the analytical and experimental results for the two-dimensional Uniform, Doubly-exponential and Gaussian and Rayleigh distributions, and also clearly specify the way by which one should extend the results for higher dimensions.

Keywords: Classification using Order Statistics (OS), Moments of OS.

1 Introduction

Pattern Recognition (PR) is the process by which unknown feature vectors are categorized into groups or classes based on their features [2]. The basic concept of traditional *parametric* PR is to model the classes based on the assumptions related to the underlying class *distributions*, and this has been achieved by performing a learning phase in which the moments of the respective classes are evaluated. However, there are some families of indicators (or distinguishing quantifiers) that have noticeably been *uninvestigated* in the Pattern Recognition (PR) literature. In particular, we refer to the use of phenomena that have utilized the properties of the *Order Statistics*

A. Thomas · B. John Oommen

School of Computer Science, Carleton University, Ottawa, Canada : K1S 5B6

* *Chancellor's Professor ; Fellow: IEEE and Fellow: IAPR.* This author is also an *Adjunct Professor* with the University of Agder in Grimstad, Norway. The work of this author was partially supported by NSERC, the Natural Sciences and Engineering Research Council of Canada.

(OS), which have not reported earlier. In [5] and [6], we proposed the paradigm named CMOS, i.e., Classification by the Moments of Order Statistics, in which the concept of OS was applied in a pioneering manner to the field of PR so as to achieve optimal (or near-optimal) accuracies for various classification problems. In this paper, we generalize those results for multi-dimensional distributions by proposing a strategy that is analogous to a Naïve-Bayes' approach, although it, really, is of an *anti-Naïve-Bayes'* paradigm. Using such a Naïve-Bayes' approach, we demonstrate how a CMOS classifier can be both designed and implemented. In order to prove our claims, we provide analytical and experimental results for the two-dimensional and multi-dimensional Uniform, Doubly-exponential, Gaussian and Rayleigh distributions, whence we show that the results are clearly conclusive.

Contributions of this Paper: The novel contributions of this paper are:

- We extend the OS-based PR for multi-dimensional distributions;
- We prove that the new approach attains the optimal bound for symmetric distributions, and a near-optimal accuracy for asymmetric distributions;
- To justify these claims, we also submit a formal analysis and the experimental results for a few distributions within the exponential family.

2 Relevant Background Areas Regarding OS

The entire theory of OS-based PR is based on the stochastic properties of the *distribution* of the OS of any random variable. Since the distribution of the OS is, in general, not easily computable, the literature rather concentrates on its moments – which, in actuality, is quite distinct from the moments of the random variable itself. The explicit form of these moments is crucial to our study and so it is presented here, in all brevity. Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be a univariate random sample of size n that follows a continuous distribution function Φ , where the probability density function (pdf) is $\varphi(\cdot)$. Let $\mathbf{x}_{1,n}, \mathbf{x}_{2,n}, \dots, \mathbf{x}_{n,n}$ be the corresponding OS. The r^{th} OS, $\mathbf{x}_{r,n}$, of the set is the r^{th} smallest value among the given random variables. The pdf of $\mathbf{y} = \mathbf{x}_{r,n}$ is:

$$f_{\mathbf{y}}(y) = \frac{n!}{(r-1)!(n-r)!} \{\Phi(y)\}^{r-1} \{1 - \Phi(y)\}^{n-r} \varphi(y),$$

where $r = 1, 2, \dots, n$. The fundamental theorem concerning the OS that we invoke is found in many papers [3, 4, 8], and is summarized below.

Let $n \geq r \geq k+1 \geq 2$ be integers. Then, since Φ is a nondecreasing and right-continuous function from $\mathbb{R} \rightarrow \mathbb{R}$, $\Phi(\mathbf{x}_{r,n})$ is uniform in $[0,1]$. If we now take the k^{th} moment of $\Phi(\mathbf{x}_{r,n})$, it has the form [4]:

$$E[\Phi^k(\mathbf{x}_{r,n})] = \frac{B(r+k, n-r+1)}{B(r, n-r+1)} = \frac{n! (r+k-1)!}{(n+k)! (r-1)!}, \quad (1)$$

where $B(a,b)$ denotes the *Beta* function, and $B(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$ since its parameters are integers.

The above fundamental result can also be used for characterization purposes as explained in [4]. If $n = 2$, implying that only *two* samples are drawn from \mathbf{x} , we can deduce from Eq. (1) that $E[\Phi^1(\mathbf{x}_{1,2})] = \frac{1}{3}, \Rightarrow E[\mathbf{x}_{1,2}] = \Phi^{-1}(\frac{1}{3})$ and $E[\Phi^1(\mathbf{x}_{2,2})] = \frac{2}{3}, \Rightarrow E[\mathbf{x}_{2,2}] = \Phi^{-1}(\frac{2}{3})$. Thus, from a computational perspective, the first moment of the first and second 2-order OS would be the values where the cumulative distribution Φ equal $\frac{1}{3}$ and $\frac{2}{3}$ respectively.

3 Uni-dimensional OS-Based Classification: The Generic Classifier

The multi-dimensional OS-based classifier is based on its uni-dimensional counterpart developed in [5] and [6]. Since the understanding of the latter is crucial to this paper, this is briefly explained here.

Consider a 2-class problem with classes ω_1 and ω_2 , where their class-conditional densities are $f_1(x)$ and $f_2(x)$ respectively. If we perform a classification based on v_1 and v_2 , the *medians* of the distributions, this is equivalent to the strategy in which the task is performed based on a *single* OS. For all symmetric distributions, this classification attains the Bayes' accuracy – which is not too astonishing because the median is identical to the mean. But the intriguing aspect emerges when we use higher order OS that are not located centrally, but rather *distant* from the means. Indeed, for uni-dimensional OS-based PR, our methodology is based on considering the n -order OSs, and comparing the testing sample with the $n - k^{th}$ OS of the first distribution and the k^{th} OS of the second. By considering the entire spectrum of the possible values of k , the results in [5] and [6] showed that the specific value of k is usually not so crucial. Further, if these symmetric pairs of the OS are used in PR, the classification based on *these* attains the optimal Bayes' bound for a large number of symmetric distributions of the exponential family [5] and a near-optimal bound for the asymmetric distributions [6].

Our task is to now generalize these results for multi-dimensional distributions. As mentioned earlier, the generalization of CMOS for multi-dimensional classification problems is not so trivial. We have opted to do this by invoking a Naïve-Bayes' approach, which essentially implies that the first moments of the OS in each of the dimensions are uncorrelated².

4 Uniform Distribution

We first extend the uni-dimensional results of [5] to show that CMOS can also attain a similar optimal bound for their multi-dimensional counterparts. To prove this claim, we first consider two-dimensional distributions, and thereafter, extend the result for higher dimensions.

² Although the uncorrelation is *sufficient*, we are not certain whether the *independence* of the features is *necessary*. As far as we are concerned, this is still an open issue.

Order Statistics: For a *prima facie* case, we consider two 2-dimensional uniform distributions U_1 and U_2 in which both the features are in $[0, 1]^2$ and $[h, 1+h]^2$ respectively. Consequently, we see that the overlapping region of the distributions forms a square. In this case, it is easy to verify that the Bayesian classifier is the diagonal that passes through the intersection points of the distributions. For the classification based on the moments of the 2-OS, because the features are independent for both dimensions, we can show that this is equivalent to utilizing the OS at position $\frac{2}{3}$ of the first distribution for both dimensions, and the OS at the position $h + \frac{1}{3}$ of the second distribution for both dimensions.

Theoretical Analysis - 2-OS: Consider the points $\mathbf{u}_1 = (\frac{2}{3}, \frac{2}{3})$ of the first distribution and $\mathbf{u}_2 = (h + \frac{1}{3}, h + \frac{1}{3})$ of the second distribution in the two-dimensional space. We now show that if we compare the testing point $\mathbf{x} = (x_1, x_2)$ with these points, the PR attains the optimal classification, i.e., that which is the result of comparing it with the corresponding means.

Theorem 1. *For the 2-class problem in which the two 2-dimensional class conditional distributions are Uniform and identical, CMOS attains the optimal Bayes' bound.*

Proof. The proof is omitted here in the interest of brevity, but is in [7]. \square

Experimental Results - 2-OS: The CMOS method for 2-dimensional uniform distributions U_1 (in $[0, 1]$ in both dimensions) and U_2 (in $[h, 1+h]$ in both dimensions) has been tested, and the results are given in Table 1. For each of the experiments, we generated 1,000 points for the classes ω_1 and ω_2 .

Table 1 Classification of Uniformly distributed 2-dimensional classes by the CMOS 2-OS method for different values of h

h	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60
Bayesian	99.845	99.505	98.875	98.045	97.15	95.555	94.14	91.82
CMOS	99.845	99.505	98.875	98.045	97.15	95.555	94.14	91.82

Theoretical Analysis - k -OS: We shall now discuss the efficiency of the k -OS CMOS. As in the case of the 2-OS CMOS, by enforcing the fact that the features are independent, we can, indeed, prove that the k -OS CMOS also attains the optimal Bayesian bound.

Theorem 2. *For the 2-class problem in which the two class conditional distributions are Uniform and identical as $U(0, 1)^2$ and $U(h, 1+h)^2$, optimal Bayesian classification can be achieved by using symmetric pairs of the n -OS, i.e., the $n-k$ OS for ω_1 (represented by u_1) and the k OS for ω_2 (represented by u_2) if and only if $k > \frac{(n+1)(1-h)}{2}$ for both the features. If this condition is violated, the CMOS classifier uses the Dual condition, i.e., the k OS for ω_1 and the $n-k$ OS for ω_2 for both the features.*

Proof. The proof can be found in [7] and is omitted here. \square

Experimental Results - k -OS: The k -OS CMOS method for 2-dimensional Uniform distributions U_1 (in $[0, 1]$ in both dimensions) and U_2 (in $[h, 1+h]$ in both dimensions) has been tested, and the results are given in Table 2.

Table 2 Classification of Uniformly distributed 2-dimensional classes by the k -OS CMOS method for different values of h . The scenarios when we have invoked the *Dual* condition are specified by the notation “(D)”

$h \longrightarrow$	0.95	0.90	0.85	0.80	0.75	0.70	0.65
$\langle \frac{2}{3}, \frac{1}{3} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73
$\langle \frac{4}{5}, \frac{1}{5} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73
$\langle \frac{6}{7}, \frac{1}{7} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69 (D)	93.73 (D)
$\langle \frac{4}{7}, \frac{3}{7} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73
$\langle \frac{8}{9}, \frac{1}{9} \rangle$	99.92	99.58	98.86	97.94	96.78 (D)	95.69 (D)	93.73 (D)
$\langle \frac{7}{9}, \frac{2}{9} \rangle$	99.92	99.58	98.86	97.94	96.78	95.69	93.73

If we examine Table 2, we can see that k -OS CMOS attained the optimal Bayes' bound for all the cases where the condition is strictly enforced. But, for the cases where the condition failed, the dual condition holds and so the CMOS positions should be reversed so as to attain the optimal accuracy. For example, consider the classification with the CMOS positions, $\langle \frac{8}{9}, \frac{1}{9} \rangle$ for $h = 0.75$. As stated earlier, for any symmetric pair, the condition which is to be enforced is that the $\frac{n-1+k}{n+1}^{th}$ percentile should be less than the $\frac{k}{n+1}^{th}$ percentile for every dimension. But, for the pairs $\langle \frac{8}{9}, \frac{1}{9} \rangle$, this is not true, and hence, the dual CMOS has to be invoked to obtain the optimal bound.

Multi-dimensional Extension: As the multi-dimensional distribution naturally imposes the independence for the Uniform scenario, we can extend the result of Theorems 1 and 2 to obtain a classifier for such a problem.

Theorem 3. *For the 2-class problem in which the two class conditional distributions are Uniform and identical as $U(0, 1)^d$ and $U(h, 1+h)^d$, the classifier $x_1 + x_2 + \dots + x_d \leq \frac{\omega_1}{\omega_2} \frac{d}{2}(h+1)$ obtained by using symmetric pairs of the n -OS, i.e., the $n-k$ OS for ω_1 and k OS for ω_2 , leads to an optimal Bayesian classification if and only if $k > \frac{(n+1)(1-h)}{2}$ for all the features. If this condition is violated, the CMOS classifier uses the Dual condition, i.e., the k OS for ω_1 and the $n-k$ OS for ω_2 for both the features.*

Proof. The proofs for higher order multi-dimensional distributions follow due to the independence and due to the arguments analogous to those used in Theorems 1 and 2. The details are omitted here in the interest of brevity. \square

5 Doubly Exponential Distribution

We earlier worked with the 2-class problem in which the class conditional distributions are uni-dimensional Doubly Exponential and identical, and demonstrated that the optimal Bayesian classification can be achieved by using symmetric pairs of the n -OS, i.e., the $n - k$ OS for the first distribution and the k -OS for the second distribution if and only if $\ln\left(\frac{2k}{n+1}\right) > \frac{c_1 - c_2}{2}$ where c_1 and c_2 are the respective means of the distributions. Now, we shall extend this result for the multi-dimensional Doubly Exponential distributions.

Order Statistics: Let ω_1 and ω_2 be the two classes where the features follow two-dimensional Doubly Exponential distributions. Then, since the random vectors have independent components³, the pdfs can be represented as:

$$f_1(\mathbf{x}) = \frac{\lambda_{11}}{2} e^{-\lambda_{11}|x_1 - c_{11}|} \cdot \frac{\lambda_{12}}{2} e^{-\lambda_{12}|x_2 - c_{12}|}, \quad -\infty < x_1 < \infty, -\infty < x_2 < \infty,$$

$$f_2(\mathbf{x}) = \frac{\lambda_{21}}{2} e^{-\lambda_{21}|x_1 - c_{21}|} \cdot \frac{\lambda_{22}}{2} e^{-\lambda_{22}|x_2 - c_{22}|}, \quad -\infty < x_1 < \infty, -\infty < x_2 < \infty,$$

where $\mathbf{c}_1 = (c_{11}, c_{12})$ and $\mathbf{c}_2 = (c_{21}, c_{22})$ are the respective means of the distributions, and the values $\lambda_{11}, \lambda_{12}, \lambda_{21}$ and λ_{22} are the corresponding parameters of the distributions in the respective dimensions.

In [5], we had derived the k -OS CMOS positions for the uni-dimensional Doubly Exponential distribution as $u_1 = c_1 - \frac{1}{\lambda_1} \ln\left(\frac{2k}{n+1}\right)$ and $u_2 = c_2 + \frac{1}{\lambda_2} \ln\left(\frac{2k}{n+1}\right)$. As the individual features of the Doubly Exponential distribution are independent, the CMOS positions are computed directly using *these* independent univariate distributions, and thus have the corresponding forms as those of the positions obtained for the uni-dimensional distributions. Consequently, for the two dimensional distributions for classes ω_1 and ω_2 , the CMOS positions \mathbf{u}_1 and \mathbf{u}_2 are $\left[c_{11} - \frac{1}{\lambda_{11}} \ln\left(\frac{2k}{n+1}\right), c_{12} - \frac{1}{\lambda_{12}} \ln\left(\frac{2k}{n+1}\right)\right]^T$ and $\left[c_{21} + \frac{1}{\lambda_{21}} \ln\left(\frac{2k}{n+1}\right), c_{22} + \frac{1}{\lambda_{22}} \ln\left(\frac{2k}{n+1}\right)\right]^T$ respectively.

Theoretical Analysis: We prove the optimal properties of CMOS for two-dimensional identical and symmetrically placed Doubly Exponential distributions, with means $(0, 0)$ and (c, c) respectively, and with identical λ .

Theorem 4. *For the 2-class problem in which the two class conditional distributions are two-dimensional Doubly Exponential, identical and symmetric, optimal Bayes' classification can be achieved by using symmetric pairs of the n-OS, i.e., the n - k OS for ω_1 and the k OS for ω_2 if and only if $\ln\left(\frac{2k}{n+1}\right) > \frac{c_1 - c_2}{2}$ for both the features. If this condition is violated, the CMOS classifier uses the Dual condition, i.e., the k OS for ω_1 and the n - k OS for ω_2 for both the features.*

Proof. The proof is not omitted here, but is included in [7]. □

³ This independence is a consequence of the fact that the exponential terms can be factored so that each factor only possesses a *single* variable.

Experimental Results: The CMOS classifier has been rigorously tested for a number of experiments with various distributions with means $\mathbf{c}_1 = \langle 0, 0 \rangle$ and $\mathbf{c}_2 = \langle c, c \rangle$ respectively. The test results are depicted in Table 3.

Table 3 Classification of Doubly Exponentially distributed 2-dimensional classes by the CMOS k -OS method for different means

No.	c	w.r.t Mean	$\langle \frac{2}{3}, \frac{1}{3} \rangle$	$\langle \frac{4}{5}, \frac{1}{5} \rangle$	$\langle \frac{5}{7}, \frac{2}{7} \rangle$	$\langle \frac{8}{9}, \frac{1}{9} \rangle$
1	3	96.55	96.55	96.55	96.55	96.55
2	2.5	95.5	95.5	95.5	95.5	95.5
3	2	92	92	92	92	92
4	1.5	89.3	89.3	89.3 (D)	89.3	89.3 (D)

Now, consider the results presented in the row denoted by Trial No. 4. In this case, the testing attained the Bayes' accuracy for the symmetric OS pairs $\langle \frac{2}{3}, \frac{1}{3} \rangle$ and $\langle \frac{5}{7}, \frac{2}{7} \rangle$, but the dual pairs had to be used for the pairs $\langle \frac{4}{5}, \frac{1}{5} \rangle$ and $\langle \frac{8}{9}, \frac{1}{9} \rangle$, since these values violated the condition imposed by Theorem 4.

Multi-Dimensional Extension: As the features are again independent because of the explicit factorizability, we can extend the result of Theorem 4 to perform a classification with respect to the $\langle \frac{n+1-k}{n+1}, \frac{k}{n+1} \rangle$ positions of each of the features, for identical and symmetrical distributions.

Theorem 5. *For the 2-class problem in which the two class conditional distributions are d -dimensional Doubly Exponential, identical and symmetric, the optimal Bayesian classifier is $x_1 + x_2 + \dots + x_d = \frac{d}{2} \cdot c$ and is exactly the CMOS classifier obtained by using symmetric pairs of the n -OS, i.e., the $n - k$ OS for ω_1 and the k OS for ω_2 , if and only if $\ln\left(\frac{2k}{n+1}\right) > \frac{c_1 - c_2}{2}$ for all the features. If this condition is violated, the CMOS classifier uses the Dual condition, i.e., the k OS for ω_1 and the $n - k$ OS for ω_2 for all the features.*

Proof. The details of the proof are omitted here to avoid repetition. \square

6 Gaussian Distribution

In this section, we intend to work with multi-dimensional Gaussian distribution. Earlier in [5], we showed that CMOS can attain optimal classification for unidimensional Gaussian distribution. The expected values of the first moment of the 2-OS can be determined as $E[\mathbf{x}_{1,2}] = \mu - \frac{\sigma}{\sqrt{2\pi}}$ and $E[\mathbf{x}_{2,2}] = \mu + \frac{\sigma}{\sqrt{2\pi}}$ as shown in [1]. We initially deal with the two-dimensional Gaussian distribution, and then extend the result for higher dimensions.

Order Statistics: Let ω_1 and ω_2 be the two classes where the features follow two-dimensional Gaussian distributions. As the pdfs of the Gaussian distributions are

not factorizable, we need to assume the independence (i.e., uncorrelation) of the features. Then, the pdfs can be represented as:

$$f_1(\mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma_{11}} e^{-\frac{(x_1-\mu_{11})^2}{2\sigma_{11}^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_{12}} e^{-\frac{(x_2-\mu_{12})^2}{2\sigma_{12}^2}}, \quad -\infty < x_1, x_2 < \infty, \text{ and}$$

$$f_2(\mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma_{21}} e^{-\frac{(x_1-\mu_{21})^2}{2\sigma_{21}^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_{22}} e^{-\frac{(x_2-\mu_{22})^2}{2\sigma_{22}^2}}, \quad -\infty < x_1, x_2 < \infty,$$

where $\mu_1 = (\mu_{11}, \mu_{12})$ and $\mu_2 = (\mu_{21}, \mu_{22})$ are the respective means and $\sigma_1 = (\sigma_{11}, \sigma_{12})$ and $\sigma_2 = (\sigma_{21}, \sigma_{22})$ are the corresponding standard deviations of the distributions.

Then, the 2-OS CMOS positions for the uni-dimensional Gaussian distribution are [5] $u_1 = \mu_1 - \frac{\sigma}{\sqrt{2\pi}}$ and $u_2 = \mu_2 + \frac{\sigma}{\sqrt{2\pi}}$. As the individual features of the Gaussian distribution are independent, the CMOS positions are computed directly using *these* independent univariate distributions, and thus have the same forms as those of the positions obtained for the uni-dimensional distributions. Thus, for the two dimensional features, the CMOS positions \mathbf{u}_1 and \mathbf{u}_2 are $\left[\mu_{11} - \frac{\sigma_{11}}{\sqrt{2\pi}}, \mu_{12} - \frac{\sigma_{12}}{\sqrt{2\pi}} \right]^T$ and $\left[\mu_{21} - \frac{\sigma_{21}}{\sqrt{2\pi}}, \mu_{22} - \frac{\sigma_{22}}{\sqrt{2\pi}} \right]^T$ respectively.

Theoretical Analysis: Without loss of generality, we consider the distributions to have the means $(0, 0)$ and (μ, μ) respectively, and with identical standard deviations, σ ($\sigma_{11} = \sigma_{12} = \sigma_{21} = \sigma_{22} = \sigma$).

Theorem 6. *For the 2-class problem in which the two 2-dimensional class conditional distributions are Gaussian, identical and symmetric, the 2-OS CMOS attains the optimal Bayes' bound.*

Proof. The proof is included in [7] and is omitted here for brevity. \square

Experimental Results: The CMOS has been rigorously tested for 2-dimensional Gaussian distributions, and the results are given in Table 4.

Table 4 Classification of 2-dimensional Gaussian distributions by the CMOS 2-OS method for different means $(0, 0)$ and (μ, μ)

μ	1	1.5	2	2.5	3	3.5	4	4.5
Bayesian	75.985	85.485	91.93	96.13	98.335	99.34	99.81	99.95
CMOS	75.985	85.485	91.93	96.13	98.335	99.34	99.81	99.95

Multi-dimensional Extension: The result of Theorem 4 can be generalized for higher dimensions. As the features are assumed to be independent, the classification can be done with regard to the 2-OS CMOS positions of each of the features for the identical and symmetrical distributions.

Theorem 7. For the 2-class problem in which the two class conditional distributions are d -dimensional Gaussian, identical and symmetric, the optimal Bayesian classifier has the form $x_1 + x_2 + \dots + x_d = \frac{d}{2} \cdot \mu$, and this is again the classifier obtained by using symmetric 2-OS CMOS positions.

Proof. The proof is straightforward and can be found in [7]. \square

7 Rayleigh Distribution

In [6], we had earlier worked with uni-dimensional Rayleigh distributions in which CMOS attained near-optimal classification with regard to the classifiers based on the medians of the distribution. We also showed that the error difference created by the CMOS classifier when compared to the Bayesian classifier is negligible by considering the differences of the error probabilities quantified by the differences between the areas under the curves of the resulting errors. We now investigate the two-dimensional scenario.

Order Statistics: Let ω_1 and ω_2 be the two classes where the features follow two-dimensional Rayleigh distributions. Earlier, in [5], we derived the 2-OS CMOS positions for the uni-dimensional Rayleigh distribution as $u_1 = \sigma\sqrt{2\ln(3)}$ and $u_2 = \theta + \sigma\sqrt{2\ln(\frac{3}{2})}$. In order to extend this result for a two-dimensional case, as before, we assume a Naïve-Bayes' approach, in which the first moments of the OS in each of the dimensions are uncorrelated. Consequently, for the two dimensional distributions for the classes the CMOS positions \mathbf{u}_1 and \mathbf{u}_2 are respectively $\mathbf{u}_1 = [\sigma_{11}\sqrt{2\ln(3)}, \sigma_{12}\sqrt{2\ln(3)}]^T$ and $\mathbf{u}_2 = [\theta_1 + \sigma_{21}\sqrt{2\ln(\frac{3}{2})}, \theta_2 + \sigma_{22}\sqrt{2\ln(\frac{3}{2})}]^T$.

Theoretical Analysis: As in the uni-dimensional case in [6], we can compute the differences in the corresponding analogous volumes created by the classifiers for the respective distributions. The “ceiling” of the volume is rather complex because it involves the difference between the corresponding three-dimensional surfaces. However, we can easily obtain an upper bound for this volume by considering the smallest bounding rectanguloid.

The maximum bound of the error can be numerically evaluated and can be found to be nearly zero. Thus, we conclude that the maximum error of the CMOS classifier is negligible and attains a near-optimal bound.

Experimental Results: The CMOS method has been rigorously tested with different possibilities of the k -OS and for various values of n , and the test results are given in Table 5.

Table 5 Classification of Rayleigh distributed 2-dimensional classes by the CMOS k -OS method for different values of θ

No.	Order(n)	Moments	$\theta = 2$	$\theta = 1.5$	$\theta = 1.3$	$\theta = 1.2$
1	Median	$\left[\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}\right)\right]$	98.25	95.3	93.15	89.75
2	Two	$\left[\left(\frac{2}{3}, \frac{1}{3}\right), \left(\frac{2}{3}, \frac{1}{3}\right)\right]$	98.35	95.3	92.95	89.6
3	Four	$\left[\left(\frac{4}{5}, \frac{1}{5}\right), \left(\frac{4}{5}, \frac{1}{5}\right)\right]$	98.65	95.3	92.7	90.6
4	Six	$\left[\left(\frac{6}{7}, \frac{1}{7}\right), \left(\frac{6}{7}, \frac{1}{7}\right)\right]$	98.75	95.15	92.2 (D)	90.35 (D)
5	Six	$\left[\left(\frac{4}{7}, \frac{3}{7}\right), \left(\frac{4}{7}, \frac{3}{7}\right)\right]$	98.25	95.2	93.15	89.75
6	Eight	$\left[\left(\frac{8}{9}, \frac{1}{9}\right), \left(\frac{8}{9}, \frac{1}{9}\right)\right]$	98.8	95.2 (D)	92.25 (D)	89.9 (D)
7	Eight	$\left[\left(\frac{7}{9}, \frac{2}{9}\right), \left(\frac{7}{9}, \frac{2}{9}\right)\right]$	98.65	95.3	92.85	90.45

8 Other Multi-dimensional Distributions

In Sections 4 - 7, we dealt with some of the multi-dimensional distributions of the exponential family and proved that the CMOS can attain the optimal Bayes' bound for all the symmetric distributions and a near-optimal bound for the asymmetric distribution. One can see that the strategy is analogous to a Naïve-Bayes' approach, although it, really, is of an *anti-Naïve-Bayes'* paradigm. For the distributions of which the class-conditional densities are not factorizable, it is important to assume that the features are uncorrelated. The same argument can be used for any higher-order distributions.

9 Conclusions

In this paper, we generalized the results of CMOS for multi-dimensional distributions. We provided the analytical and experimental results for the two-dimensional Uniform, Doubly-exponential, Gaussian and Rayleigh distributions, and also generalized the approach for higher dimensions. We have showed that CMOS can attain the optimal Bayes' bound for symmetric distributions and near-optimal results for asymmetric distributions. The analogous results for the other distributions in the exponential family, which were discussed in [5, 6] are also available, but omitted here to avoid repetition.

References

1. Ahsanullah, M., Nevzorov, V.B.: Order Statistics: Examples and Exercises. Nova Science Publishers, Inc. (2005)
2. Duda, R.O., Hart, P.: Pattern Classification and Scene Analysis. A Wiley Interscience Publication (2000)

3. Morris, K.W., Szynal, D.: A goodness-of-fit for the Uniform Distribution based on a Characterization. *Journal of Mathematical Science* 106, 2719–2724 (2001)
4. Lin, G.D.: Characterizations of Continuous Distributions via Expected values of two functions of Order Statistics. *Sankhya: The Indian Journal of Statistics* 52, 84–90 (1990)
5. Thomas, A., Oommen, B.J.: The Fundamental Theory of Optimal “Anti-Bayesian” Parametric Pattern Classification Using Order Statistics Criteria. *Pattern Recognition* 46, 376–388 (2013)
6. Thomas, A., Oommen, B.J.: Optimal Order Statistics-based “Anti-Bayesian” Parametric Pattern Classification for the exponential family. *Pattern Recognition* (2012)
7. Thomas, A., Oommen, B.J.: Order Statistics-based Parametric Classification for Multi-dimensional Distributions (2012) (to be submitted)
8. Too, Y., Lin, G.D.: Characterizations of Uniform and Exponential Distributions. *Academia Sinica* 7(5), 357–359 (1989)

Knowledge Extraction from Graph-Based Structures in Conceptual Design

Grażyna Ślusarczyk

Abstract. This paper deals with using knowledge in order to support the conceptual stage of the design process. Both a visual language composed of design drawings and the internal graph-based structures representing these drawings are specified on the basis of the specified conceptualization of the design domain. The defined ontological commitment enables to transform knowledge about drawings encoded in their internal representations into logic formulas. The extracted facts concerning drawings together with axioms describing general domain-oriented design knowledge allow the design support system to reason about validity of created design solutions. The approach is illustrated on examples of designing floor layouts.

1 Introduction

Although there are many CAD tools for describing, editing, analyzing, and evaluating design projects, the conceptual design phase is the least supported one [7]. This paper deals with extracting and processing knowledge which allows the design system to support the conceptual stage of the design process. The presented method is an extension of the approach proposed in [5, 6]. It describes early design drawings with internal graph-based representations created automatically, and proposes a logic-based method of reasoning about validity of projects. A clear distinction between axioms describing design constraints which must be obeyed and formulas describing requirements specified by the user is made.

In this paper a conceptualization of a visual design domain, which specifies concepts that are assumed to exist in this domain and relationships that hold among them [8] is outlined. This conceptualization constitutes the basis for defining a

Grażyna Ślusarczyk

The Faculty of Physics, Astronomy and Applied Computer Science

Jagiellonian University, Reymonta 4, 30-059 Kraków, Poland

e-mail: gslusarc@uj.edu.pl

problem-oriented visual language composed of design drawings, which contain general ideas about design objects, and is related to the internal representation of drawings in the form of hypergraphs, where hypergraph atoms represent concepts and relations of the domain.

Hypergraphs encode the syntactic and semantic information about early design solutions [8]. Their hyperedges represent drawing components and multi-argument relations among fragments of components. Due to the specified ontological commitment between elements of the vocabulary of the first-order logic language and entities of the conceptualization the knowledge stored in hypergraphs is extracted in the form of atomic formulas describing designs.

The reasoning module of the proposed system tests the consistency of drawings with axioms describing general domain-specific design knowledge and requirements/constraints expressed as first-order logic formulas by means of inference on the basis of atomic formulas (facts) describing designs. The inference reasoning, where the validity of logic formulas is checked using the software package ANTLR (ANother Tool for Language Recognition), is discussed. The resulting assessment of drawings given by the system supports decision-making process throughout the whole design process.

The approach is illustrated on examples of designing floor layouts created by means of our prototype HSSDR system [6].

2 Visual Design System

A schema of the proposed visual design system (called HSSDR) is shown in Fig.1. The design interface allows the designer to create design drawings by means of a problem-oriented visual language. A rule editor being a part of the design interface enables to define design constraints and requirements which should be obeyed during the whole design process.

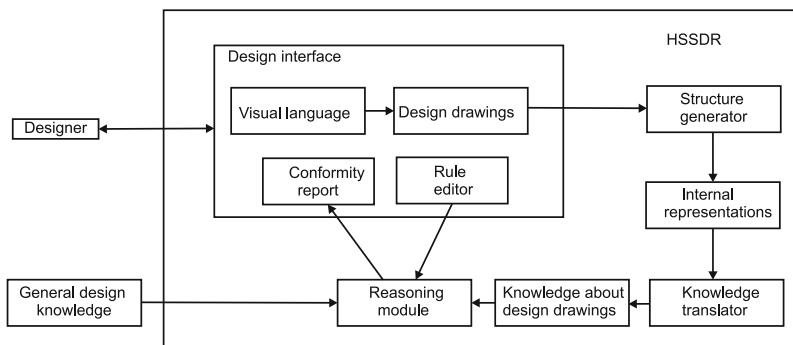


Fig. 1 A schema of the proposed computer-aided visual design system

The design drawings are automatically transformed into their internal representations in the form of hypergraphs by the structure generator module. The design knowledge stored in these structures is transformed into atomic formulas (facts) describing design drawings by the knowledge translator module. Then, the reasoning module checks the validity of design solutions by comparing logic formulas expressing both general design knowledge specific to a particular design task and design requirements and/or constraints with facts concerning created designs. The module makes the conformity report, which is treated as a feedback provided to the designer. The system supports rapid and intelligent decision-making throughout the whole conceptual design process.

3 Domain-Specific Visual Design Language

During the initial phase of design the conceptualization of the design domain being a classification and categorization of the knowledge [1,3] concerning this domain is defined. It specifies domain concepts, their attributes, taxonomy and relations which can hold among concepts. Each specialized design domain has its own drawing components related to concepts used in this domain and a convention which allows to express connections among drawing components according to relations of the conceptualization. The first level of the ontological commitment in the man-machine interaction is specified between mental images of design concepts and relations, and their externalization in the form of drawing components and their arrangements. Admissible layouts of drawing components in the specified domain form a specialized visual language.

The designer communicates with the design system using a visual editor which enables him to create design drawings with the use of a domain-specific visual language. This language is characterized by a vocabulary being a finite set of basic drawing components and a finite set of rules specifying possible configurations of these components.

Let us assume that a design task is to create a layout of one floor of an office building. On the basis of general requirements concerning this task the designer generates a design drawing visualizing an early solution (Fig. 2). A vocabulary of the visual language used in this case is composed of shapes corresponding to design concepts being rooms, walls, doors and motion sensors, i.e., polygons, line segments, small rectangles and circles, respectively. Thus the design drawing is composed of polygons which are placed in an orthogonal grid and represent rooms of a floor layout. The adjacency and accessibility between rooms, and fastening of sensors on walls constitute the considered relations between design concepts. The adjacency relation between rooms is expressed by line segments shared by polygons, while the accessibility relation is represented by line segments with small rectangles located on them. Black dots located on line segments represent sensors, while grey circle sectors correspond to spatial ranges of sensors. It should be noted that range spaces of sensors (the region of space that lies within their scope) do not correspond

to any physical entities nevertheless they are treated as drawing elements [2]. The sectors are not drawn by the designer, but they are automatically generated by the system on the basis of the sensor attributes specified by the designer while placing sensors in the drawing. Thus, the designer has the possibility to model constraints involving spatial functionality of objects.

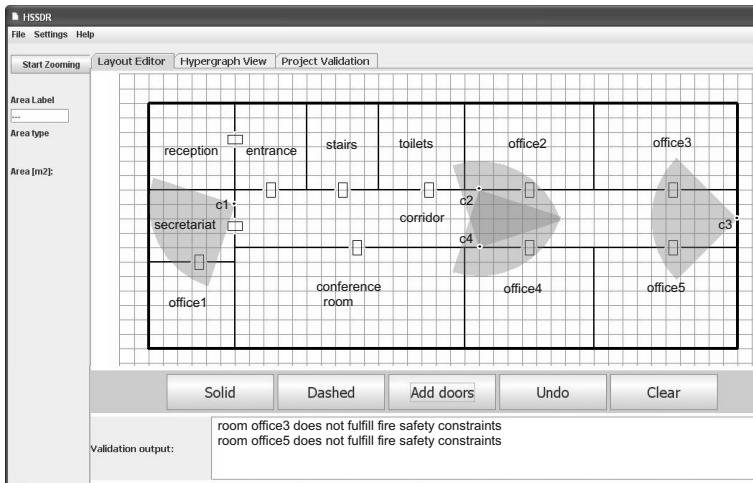


Fig. 2 A design of a floor layout of an office building with four sensors

4 Graph-Based Data Structures

In the considered system, the design drawings have internal representations in the form of attributed hypergraphs [5]. Hypergraph atoms represent concepts and relations corresponding to the elements of the conceptualization determined by the designer. Hypergraphs enable the system to store knowledge about syntactic and semantic aspects of created drawings.

The proposed hypergraphs have two types of hyperedges, called *object hyperedges* and *relational hyperedges*. Hyperedges of the first type correspond to design components and are labelled by component names. Hyperedges of the second type represent relations among fragments of components and can be either directed or non-directed. They are labelled by names of relations. Object hyperedges are connected with relational hyperedges by means of hypergraph nodes corresponding to fragments of design components. Attributes assigned to hyperedges and nodes represent properties of the corresponding design components and relations between them.

A hypergraph corresponding to the floor layout presented in Fig.2 is shown in Fig.3. It contains 16 object hyperedges, e_1, \dots, e_{16} , (denoted by rectangles), where 12 of them represent rooms and 4 represent sensors. They are connected with the

relational ones (denoted by ovals) by means of hypergraph nodes which represent walls of rooms or fragments of sensors. The numbers assigned to hypergraph nodes representing walls correspond to the numbers of polygon sides, which are ordered clock-wise. There are 11 relational hyperedges (labelled *acc*) representing the accessibility relation between rooms, 8 relational hyperedges (labelled *adj*) representing the adjacency relation, and 4 directed relational hyperedges (labelled *on*) representing sensor fastening.

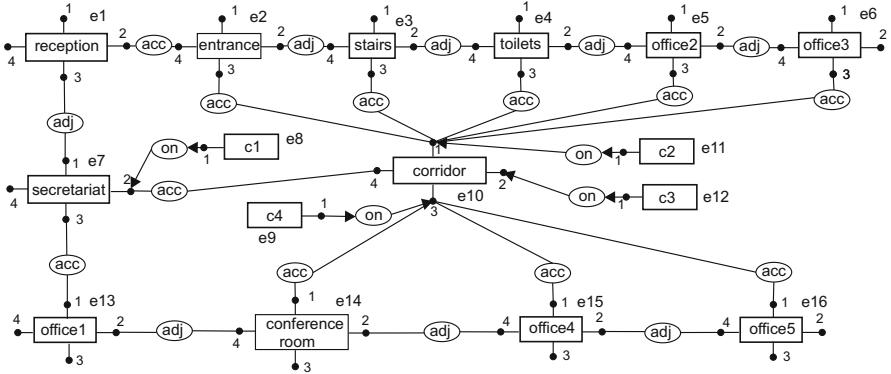


Fig. 3 A hypergraph corresponding to the floor layout presented in Fig.2

The values of such attributes as *area*, which are assigned to object hyperedges representing rooms, are automatically set by the system at the time of creating rooms on the basis of the occupied part of the grid on which designs are drawn. The values of other attributes, like *material* characterizing walls or *range* characterizing sensors, are specified by the designer at the time of establishing the relations between components. Attributes *length*, *door_number* and *loc_door* assigned to nodes representing walls of rooms specify the length of a wall, the number of doors in a wall and the location of doors, respectively.

5 Knowledge Extraction and Reasoning Mechanism

On the basis of the set of initial design requirements the designer externalizes his/her mental model of a design solution in the form of a design drawing. While analyzing successive drawings the designer infers useful information what results in adding the devised requirements or changing the existing ones. This process can be effectively supported by the interaction between the designer and computer during the whole design process. The visualization of solutions together with the feedback obtained from the system, which assesses the solutions, make the designer search for different possibilities of developing the solution further. He/she can either

consider new requirements or apply actions which lead to the fulfillment of still unsatisfied ones.

In the proposed approach information stored in hypergraphs corresponding to design drawings is translated to atomic formulas of the first-order logic, called *facts*. The reasoning module checks validity of design drawings by comparing these facts with first-order logic formulas expressing both general design knowledge and design requirements.

In order to define logic formulas their vocabulary $T = \{C, F, P\}$, where C is a set of constant symbols, F is a set of multi-argument function symbols, and P is a set of multi-argument predicates, is specified. Then the mapping called *ontological commitment* between elements of the vocabulary and entities of the conceptualization is defined. The drawing components are assigned to the constant symbols, their attributes are assigned to function symbols, while relations between the components correspond to the predicate symbols. This commitment allows the system to transform semantic and syntactic information encoded in the internal representations of drawings into logic formulas.

While designing floor layouts elements of a set of conceptualization concepts (rooms, walls, sensors) and relations (adjacency, accessibility and fastening) are associated with symbols of the vocabulary of the logic language. Walls of rooms, rooms and sensors are associated with constant symbols, while relations between them are assigned to predicate symbols. Attributes determined for walls, rooms and sensors correspond to function symbols.

We assume that we have a set of variables. The set of *terms* is formed starting from constant symbols and variables and closing off under function application, i.e., if t_1, \dots, t_n , $n \geq 1$, are terms and $f \in F$ is a n -ary function symbol, then $f(t_1, \dots, t_n)$ is also a term. An *atomic formula* is either of the form $p(t_1, \dots, t_k)$, where $p \in P$ is a k -ary relation symbol and t_1, \dots, t_k are terms, or of the form $t_1 = t_2$, where t_1 and t_2 are terms. The set of general logical formulas is built over atomic ones using logical connectives and quantifiers.

The semantics of first-order formulas uses a *relational structure* consisting of a domain of individuals and a way of associating with elements of the vocabulary corresponding entities over the domain [4]. A relational T -structure L consists of a domain denoted $\text{dom}(L)$, an assignment of a k -ary relation $p^L \subseteq \text{dom}(L)^k$ to each k -ary relation symbol $p \in P$, an assignment of a n -ary function $f^L : \text{dom}(L)^n \rightarrow \text{dom}(L)$ to a n -ary function symbol $f \in F$, and an assignment of a $c^L \in \text{dom}(L)$ to each constant symbol c .

In the proposed approach, where structures of drawings have hypergraph-based representations, the domain of the relational structure contains hypergraphs. The relational structure assigns hypergraph nodes, object hyperedges and their labels to constants and terms, their attributes to functions, and relational hyperedges to predicates of formulas. The interpretation of each predicate is a relational hyperedge coming from nodes of at least one object hyperedge and coming into nodes of other object hyperedges.

In case of designing floor layouts, the relational structure assigns nodes representing walls, object hyperedges representing rooms and sensors to terms, and their

attributes to functions. The relations considered in this case, acc , adj and on , are binary, the first two are undirected ones, while the third one is directed. Each of these relations holds among design components if in the hypergraph there exists a relational hyperedge e_r labelled by the name of relation p^L ($p^L = lb(e_r)$) which connects nodes assigned to two different object hyperedges e_1 and e_2 corresponding to the design components and labelled lab_1 , lab_2 , where $lab_1 = lb(e_1)$, $lab_2 = lb(e_2)$ and lb is a hyperedge labelling function. In this case the formula $p(lab_1, lab_2)$ is created for the considered drawing. For each attribute f^L assigned to a hypergraph node v or hyperedge e a formula $f(lab) = n$ is created, where $lab = lb(v)$ or $lab = lb(e)$, and n is an admissible value of the attribute f^L . The above two types of formulas form a set Δ of atomic formulas describing the design drawing.

The computer system evaluates design drawings on the basis of a set Δ of atomic first-order logic formulas extracted from hypergraph representations of these drawings. These atomic formulas (facts) are related to topological properties of drawings encoded in hypergraph structures, and both geometrical and non-geometrical ones, which are stored in hypergraph attributes. The reasoning module automatically checks if the facts concerning the generated drawing are consistent with the axioms being first-order logic formulas expressing general design knowledge specific to a particular design task. The set of axioms Φ describes design standards like architectural norms, fire regulations, etc. In the next step facts describing drawings are compared with design constraints and requirements, which are also in the form of logic formulas of a set Ψ . Moreover, there exists the possibility to specify designer's own requirements and restrictions, which are added to the set Ψ , using a rule editor being a part of the design interface.

In order to check if a given formula is true in a structure L a specification of an interpretation of variables is needed. A valuation ω on a structure L is a function from variables to elements of $dom(L)$. Given a structure L , a valuation ω on L is inductively extended to functions that maps terms to elements of $dom(L)$.

Let us assume that for a given design drawing d , $dom(L)$ contains the hypergraph representing d . A design drawing d satisfies a formula $\varphi \in (\Phi \cup \Psi)$ if there exists a valuation ω from terms of φ to elements of $dom(L)$ such that φ is true under this valuation in L ($(L, \omega) \models \varphi$). It means that φ corresponds to a formula δ build over the set Δ of atomic formulas describing drawing d .

The reasoning module of the system in order to check if the design drawing satisfies a given formula φ decomposes this formula for each valuation of its terms into atomic parts which are compared with facts of Δ extracted from the hypergraph representation of this drawing. After checking all formulas of $\Phi \cup \Psi$ the system presents the report through the design interface.

Let us consider designing of the layout of a floor of an office building. Some of the design requirements can be as follows:

- $\psi_1 \equiv \exists t_1, t_2, t_3, t_4 : t_i = office_i$, i.e., there should be at least four offices on a floor,
- $\psi_2 \equiv \exists t_1, t_2 : t_1 = secretariat, t_2 = office_i \wedge acc(t_1, t_2)$, i.e., at least one office should be accessible from a secretariat,
- $\psi_3 \equiv \forall t_i = office_i \exists t_j \in sensor : observed(t_i, t_j)$, i.e., doors to all offices should be monitored by sensors.

Automatically obtained atomic formulas describing relations which hold among layout elements of a designed floor (Fig. 2) concern adjacency and accessibility between rooms, and fastening sensors on room walls. For example the relations concerning the room labelled *secretariat* presented in Fig. 2 are described by the following atomic formulas:

- $\delta_1 \equiv acc(secretariat, corridor)$ - accessibility from the corridor,
- $\delta_2 \equiv acc(secretariat, office_1)$ - accessibility from the office number 1,
- $\delta_3 \equiv adj(secretariat, reception)$ - adjacency to the reception, and
- $\delta_4 \equiv on(c_1, secretariat)$ - fastening the sensor number 1 on a wall of the secretariat.

The design solution presented in Fig. 2 satisfies requirements ψ_1, ψ_2, ψ_3 . Formula ψ_1 is satisfied as the reasoning module finds a valuation $\omega(t_1) = e_{13}, \omega(t_2) = e_5, \omega(t_3) = e_6, \omega(t_4) = e_{15}, \omega(t_5) = e_{16}$, with $lb(e_{13}) = office1, lb(e_5) = office2, lb(e_6) = office3, lb(e_{15}) = office4, lb(e_{16}) = office5$, where e_5, e_6, e_{13}, e_{15} and e_{16} are object hyperedges and lb is a hyperedge labelling function. Formula ψ_2 is satisfied when $\omega(t_1) = e_7, \omega(t_2) = e_{13}$, as $lb(e_7) = secretariat, lb(e_{13}) = office1$ and we have the formula of $\Delta \delta_2 \equiv acc(secretariat, office_1)$.

The predicate $observed(t_i, t_j)$, is satisfied if the doors of the room corresponding to term t_i are in the spatial range of the sensor corresponding to term t_j . In order to test this condition the system first represents it as a formula composed of atomic parts expressing the location of the sensor labelled t_j on the wall of one of the rooms ($on(t_j, t_k)$), the value of the attribute *range* defined for the sensor t_j ($range(t_j)$) and the value of the attribute *loc_door* assigned to the wall number w of the room t_i ($loc_door(t_i.w)$). In other words $observed(t_i, t_j) \Leftrightarrow on(t_j, t_k) \wedge inrange(range(t_j), loc_door(t_i.w))$, where the predicate *inrange*, is satisfied if the door located in the wall $t_i.w$ are in the spatial range of sensor t_j . Then the system checks if the rectangle representing the door is inside the fragment of the circle representing the range of the sensor. In Fig. 2 the doors of rooms *office1*, *office2* and *office4* are observed by sensors c_1, c_2 and c_4 , respectively, while the doors of rooms *office3* and *office5* are observed by sensor c_3 , and thus the formula ψ_3 is satisfied.

The conformity of the floor layout from Fig. 2 with the Polish Fire Code regulations is also checked. The regulations require that each evacuation route leading to a staircase should be not longer than 30 meters. The axiom corresponding to this condition for all corridor doors has the following form: $\phi \equiv \forall x \in wall, blg(x) = corridor, door_number(x) \geq 1 \exists x' \in wall, blg(x') = staircase, door_number(x') \geq 1 : \forall i \in \{1, \dots, door_number(x)\} \exists j \in \{1, \dots, door_number(x')\} : dist(loc_door_i(x), loc_door_j(x')) \leq 30$, where blg is the function specifying the room to which a given wall belongs (i.e., it assigns an object hyperedge to one of its nodes), $dist$ is a function computing the distance between two points, and the wall attribute *loc_door* specifies the coordinates of the door located on the wall. The condition is not satisfied for the doors of the *office3* and *office5*, thus the message about it is shown in the bottom panel in Fig. 2. The situation can be easily corrected by moving the doors or changing the location of the staircase.

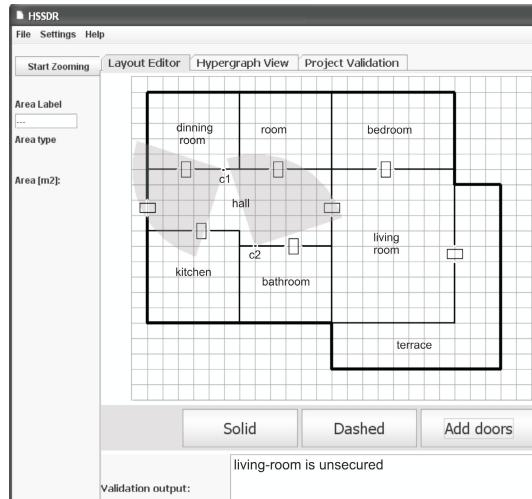


Fig. 4 A house floor layout with the unsecured living-room

After each design step the system checks which design constraints are not satisfied and shows the appropriate messages on the monitor screen. Modifications of drawings, automatically impose changes both in the hypergraph structures and facts concerning drawings which are extracted from these structures. Thus the set of atomic formulas describing designs forms dynamic knowledge about design solutions.

Let the design task be creating a floor layout of a one-storey house with monitoring which ensures the security. It means that all doors leading outside should

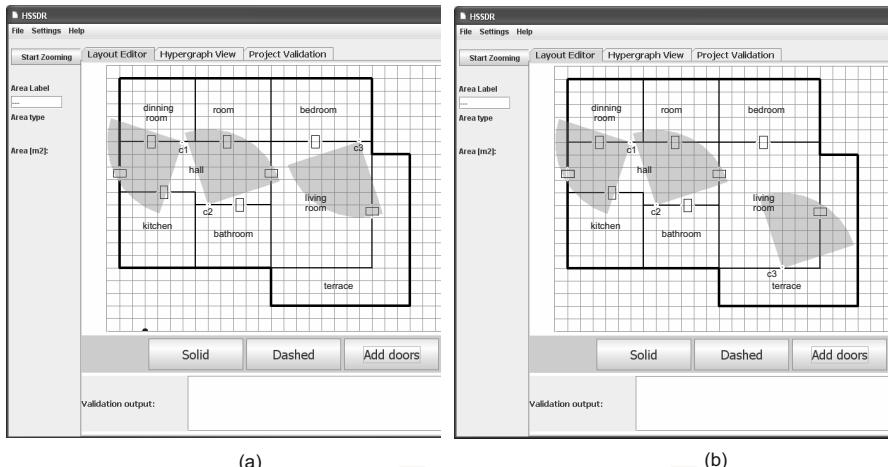


Fig. 5 A house floor layout with different placement of sensors

be observed by motion sensors. For the floor layout presented in Fig. 4 the system shows the message saying that there exist rooms which are not secured properly. The designer can find many alternative places for adding new sensors in order to satisfy design constraints (Fig. 5a and Fig. 5b).

6 Conclusions

This paper deals with using design knowledge to support the conceptual phase of design. The conceptualization of the design domain and ontological commitment are used to specify the method of automatic reasoning based on information stored in hypergraph structures corresponding to design drawings. The proposed logic model of reasoning uses atomic formulas extracted from attributed hypergraphs. Based on these formulas the verification of logic formulas representing design requirements specific to particular design tasks and defined by the designer is preformed. The consistence of design solutions with axioms expressing general design knowledge is also tested. The proposed inference method has been described on examples of designing floor layouts with the use of the prototype visual design system. Its reasoning module uses ANTLR software to verify logic formulas by decomposing them into atomic parts and comparing with formulas obtained from hypergraphs.

In the future the system will be extended by 3D visual languages allowing form-oriented design. The externalization of design ideas could be also supported by means of shape grammars. Generation methods used to create architectural forms will require using 3-D shape grammars.

References

1. Coyne, R.D., Rosenman, M.A., Radeford, A.D., Balachandran, M., Gero, J.S.: Knowledge based design systems. Addison-Wesley Publishing Company (1990)
2. Bhatt, M., Freksa, C.: Spatial computing for design: An artificial intelligence perspective. In: Visual and Spatial Reasoning for Design Creativity (SDC 2010) (2011)
3. Bhatt, M., Hois, J., Kutz, O.: Ontological modelling of form and function for architectural design. *Journal of Applied Ontology* 7, 233–267 (2012)
4. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning About Knowledge. MIT Press, Cambridge (1995)
5. Grabska, E., Lachwa, A., Ślusarczyk, G., Grzesiak-Kopeć, K., Lembas, J.: Hierarchical Layout Hypergraph Operations and Diagrammatic Reasoning. *Machine GRAPHICS and VISION* 16, 23–38 (2007)
6. Grabska, E., Borkowski, A., Palacz, W., Gajek, S.: Hypergraph system supporting design and reasoning. In: EG-ICE International Workshop, Berlin (2009)
7. Grabska, E., Ślusarczyk, G.: Knowledge and reasoning in design systems. *Automation in Construction* 22, 927–934 (2011)
8. Ślusarczyk, G.: Visual language and graph-based structures in conceptual design. *Advanced Engineering Informatics* 26, 267–279 (2012)

Estimation of the Relations of: Equivalence, Tolerance and Preference on the Basis of Pairwise Comparisons

Leszek Klukowski

Abstract. The paper presents the estimators of three relations: equivalence, tolerance and preference in a finite set on the basis of multiple pairwise comparisons, disturbed by random errors; they have been developed by the author. The estimators can rest on: binary (qualitative), multivalent (quantitative) and combined comparisons. The estimates are obtained on the basis of discrete programming tasks. The estimators require weak assumptions about distributions of comparisons errors, especially allow non-zero expected values. The estimators have good statistical properties, in particular consistency. The paper summarizes the results obtained by the author; the broader view is presented in Klukowski 2011a.

1 Introduction

Estimation of the relations of equivalence, tolerance, or preference, on the basis of multiple pairwise comparisons with random errors, is aimed at determination of an actual structure of data. It also provides the properties of estimates: consistency, distributions of errors, efficiency, etc.

The approach applied in the work rests on a statistical and optimisation paradigms: to determine the relation form, which minimizes the inconsistencies (differences) with a sample - in the form of multiple pairwise comparisons: statistical tests, experts' opinions or other procedures, prone to generating random errors. The approach presented here is an original contribution of the author to the subject. The comparisons are assumed in two forms: binary - expressing qualitative features, e.g. the direction of preference, and multivalent - expressing quantitative features of a pair, e.g. the difference of ranks of elements. The assumptions about distributions of errors of comparisons are weaker than those commonly used in the literature (David,

Leszek Klukowski

Systems Research Institute Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland
e-mail: Leszek.Klukowski@ibspan.waw.pl

1988; Slater 1961). The estimators can be applied also in the case of unknown distributions of comparison errors.

The estimators have good statistical properties, obtained on the basis of: properties of random variables expressing differences between the relation form and comparisons, the probabilistic inequalities (Hoeffding 1963, Chebyshev), properties of order statistics (David, 1970). The properties guarantee consistency of estimates for the number of independent comparisons of each pair approaching infinity.

The literature on pairwise comparisons with random errors concerns mainly ranking problems – classical results are presented in: David (1988), Bradley (1976, 1984), Davidson (1976) (bibliography), Brunk (1960). The authors mentioned present and discuss a complete range of existing methods: assumptions, estimators and their properties, tests for validation of results.

The literature concerning classification methods, based on pairs of elements is extremely extensive (see e.g. Gordon 1999, Hand 1986, Kaufman, Rousseeuw 1990, Hastie, Tibshirani, Friedman 2002, Kohonen 1995, Hartigan 1975). However, it should be emphasized that existing approaches do not cover entirely the problems presented in the work.

The paper consists with 5 sections. The second section presents main ideas of estimation, in particular the form of estimators. The next section – presents properties of estimators obtained by the author. The forth section discusses briefly optimization algorithms, which can be applied for determining of estimates. Last section summarizes results of the author in the area under consideration and shows problems for further researches.

2 Estimation of the Relations – Main Ideas

2.1 Definitions, Notations and Formulation of the Estimation Problems

The problem of estimation of relation on the basis of pairwise comparisons can be stated as follows.

We are given a finite set of elements $X = \{x_1, \dots, x_m\}$ ($3 \leq m < \infty$). There exists in the set \mathbf{X} : the equivalence relation $\mathbf{R}^{(e)}$ (reflexive, transitive, symmetric), or the tolerance relation $\mathbf{R}^{(\tau)}$ (reflexive, symmetric), or the preference relation $\mathbf{R}^{(p)}$ (alternative of the equivalence relation and strict preference relation). Each relation generates some family of subsets $\chi_1^{(\ell)*}, \dots, \chi_n^{(\ell)*}$ ($\ell \in \{p, e, \tau\}; n \geq 2$).

The equivalence relation generates the family $\chi_1^{(e)*}, \dots, \chi_n^{(e)*}$ having the following properties:

$$\bigcup_{q=1}^n \chi_q^{(e)*} = \mathbf{X}, \quad (1)$$

$$\chi_r^{(e)*} \cap \chi_s^{(e)*} = \{\mathbf{0}\} \quad (2)$$

where:

$\mathbf{0}$ – the empty set,

$$x_i, x_j \in \chi_r^{(e)*} \equiv x_i, x_j - \text{equivalent elements}, \quad (3)$$

$$(x_i \in \chi_r^{(e)*}) \cap (x_j \in \chi_s^{(e)*}) \equiv x_i, x_j - \text{non-equivalent elements for } i \neq j, r \neq s. \quad (4)$$

The tolerance relation generates the family $\chi_1^{(\tau)*}, \dots, \chi_n^{(\tau)*}$ with the property (1), i.e. $\bigcup_{q=1}^n \chi_q^{(\tau)*} = \mathbf{X}$, and the properties:

$$\exists r, s (r \neq s) \text{ such that } \chi_r^{(\tau)*} \cap \chi_s^{(\tau)*} \neq \{\mathbf{0}\},$$

$$x_i, x_j \in \chi_r^{(\tau)*} \equiv x_i, x_j - \text{equivalent elements}, \quad (5)$$

$$(x_i \in \chi_r^{(\tau)*}) \cap (x_j \in \chi_s^{(\tau)*}) \equiv x_i, x_j - \text{non-equivalent elements for } i \neq j \text{ and } (x_i, x_j) \notin \chi_r^{(\tau)*} \cap \chi_s^{(\tau)*}, \quad (6)$$

each subset $\chi_r^{(\tau)*}$ ($1 \leq r \leq n$) includes an element x_i such that

$$x_i \notin \chi_s^{(\tau)*} (s \neq r). \quad (7)$$

The preference relation generates the family $\chi_1^{(p)*}, \dots, \chi_n^{(p)*}$ with the properties (1), (2) and the property:

$$(x_i \in \chi_r^{(p)*}) \cap (x_j \in \chi_s^{(p)*}) \equiv x_i \text{ is preferred to } x_j \text{ for } r < s. \quad (8)$$

The relations defined by the conditions (1) - (8) can be expressed, alternatively, by the values (functions) $T_v^{(\ell)}(x_i, x_j)$, $((x_i, x_j) \in \mathbf{X} \times \mathbf{X}; \ell \in \{p, e, \tau\}, v \in \{b, \mu\}$; symbols b, μ denote – respectively – the binary and multivalent comparisons), defined as follows:

$$T_b^{(e)}(x_i, x_j) = \begin{cases} 0 & \text{if exists } r \text{ such that } (x_i, x_j) \in \chi_r^{(e)*}, \\ 1 & \text{otherwise;} \end{cases} \quad (9)$$

- the function $T_b^{(e)}(x_i, x_j)$, describing the equivalence relation, assuming binary values, expresses the fact if a pair (x_i, x_j) belongs to a common subset or not;

$$T_b^{(\tau)}(x_i, x_j) = \begin{cases} 0 & \text{if exists } r, s \text{ (\(r = s\) not excluded) such that} \\ & (x_i, x_j) \in \chi_r^{(\tau)*} \cap \chi_s^{(\tau)*}, \\ 1 & \text{otherwise;} \end{cases} \quad (10)$$

- the function $T_b^{(\tau)}(x_i, x_j)$, describing the tolerance relation, assuming binary values, expresses the fact if a pair (x_i, x_j) belongs to any conjunction of subsets (also to the same subset) or not; the condition (7) guarantees uniqueness of the description;

$$T_\mu^{(\tau)}(x_i, x_j) = \#(\Omega_i^* \cap \Omega_j^*), \quad (11)$$

where:

Ω_l^* – the set of the form $\Omega_l^* = \{s \mid x_l \in \chi_s^{(\tau)*}\}$,

$\#(\Xi)$ – the number of elements of the set Ξ ;

- the function $T_\mu^{(\tau)}(x_i, x_j)$, describing the tolerance relation, assuming multivalent values, expresses the number of subsets of conjunction including both elements; condition (7) guarantees the uniqueness of the description;

$$T_b^{(p)}(x_i, x_j) = \begin{cases} 0 & \text{if there exists } r \text{ such that } (x_i, x_j) \in \chi_r^{(p)*}, \\ -1 & \text{if } x_i \in \chi_r^{(p)*}, x_j \in \chi_s^{(p)*} \text{ and } r < s; \\ 1 & \text{if } x_i \in \chi_r^{(p)*}, x_j \in \chi_s^{(p)*} \text{ and } r > s; \end{cases} \quad (12)$$

- the function $T_b^{(p)}(x_i, x_j)$, describing the preference relation, assuming binary values, expresses the direction of preference in a pair or the equivalence of its elements;

$$T_\mu^{(p)}(x_i, x_j) = d_{ij} \Leftrightarrow x_i \in \chi_r^{(p)*}, x_j \in \chi_s^{(p)*}, d_{ij} = r - s; \quad (13)$$

- the function $T_\mu^{(p)}(x_i, x_j)$, describing the preference relation, assuming multivalent values, expresses the difference of ranks of elements x_i and x_j .

2.2 Assumptions about Pairwise Comparisons

The relation $\chi_1^{(\ell)*}, \dots, \chi_n^{(\ell)*}$ is to be estimated on the basis of N ($N \geq 1$) comparisons of each pair $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$; any comparison $g_{vk}^{(\ell)}(x_i, x_j)$ evaluates the actual value of $T_v^{(\ell)}(x_i, x_j)$ and can be disturbed by a random error. The following assumptions concerning the comparison errors are made:

A1. The relation type (equivalence or tolerance or preference) is known, the number of subsets n - unknown.

A2. Any comparison $g_{vk}^{(\ell)}(x_i, x_j)$ ($\ell \in \{e, \tau, p\}; v \in \{b, \mu\}; k = 1, \dots, N$), is the evaluation of the value $T_v^{(\ell)}(x_i, x_j)$, disturbed by a random error. The probabilities of errors $g_{vk}^{(\ell)}(x_i, x_j) - T_v^{(\ell)}(x_i, x_j)$ have to satisfy the following assumptions:

$$\begin{aligned} P(g_{bk}^{(\ell)}(x_i, x_j) - T_b^{(\ell)}(x_i, x_j) = 0 \mid T_b^{(\ell)}(x_i, x_j) = \kappa_{bij}^{(\ell)}) &\geq 1 - \delta \\ (\kappa_{bij}^{(\ell)} \in \{-1, 0, 1\}, \quad \delta \in (0, 1/2)), \end{aligned} \quad (14)$$

$$\begin{aligned} \sum_{r \leq 0} P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) &> 1/2 \\ (\kappa_{\mu ij}^{(\ell)} \in \{0, \dots, \pm m\}, \quad r - \text{zero or an integer number}), \end{aligned} \quad (15)$$

$$\begin{aligned} \sum_{r \geq 0} P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = -r \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) &> 1/2 \\ (\kappa_{\mu ij}^{(\ell)} \in \{0, \dots, \pm m\}, \quad r - \text{zero or an integer number}), \end{aligned} \quad (16)$$

$$P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r) \geq P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r + 1 \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) \quad (\kappa_{\mu ij}^{(\ell)} \in \{0, \dots, m\}, \quad r > 0), \quad (17)$$

$$P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r) \geq P(g_{\mu k}^{(\ell)}(x_i, x_j) - T_{\mu}^{(\ell)}(x_i, x_j) = r - 1 \mid T_{\mu}^{(\ell)}(x_i, x_j) = \kappa_{\mu ij}^{(\ell)}) \quad (\kappa_{\mu ij}^{(\ell)} \in \{0, \dots, m\}, \quad r < 0), \quad (18)$$

A3. The comparisons $g_{vk}^{(\ell)}(x_i, x_j)$ ($\ell \in \{e, \tau, p\}; v \in \{b, \mu\}$; $(x_i, x_j) \in \mathbf{X} \times \mathbf{X}$; $k = 1, \dots, N$) are independent random variables.

The assumption A3 makes it possible to determine the distributions of estimation errors of estimators. However, determination of the exact distributions of the (multidimensional) errors, in an analytic way, is complicated and, in practice, unrealizable. The main properties of the estimators, especially their consistency, are valid without the assumption.

The assumption A3 can be relaxed in the following way: the comparisons $g_{vk}^{(\ell)}(x_i, x_j)$ and $g_{vl}^{(\ell)}(x_r, x_s)$ ($l \neq k; r \neq i, j; s \neq i, j$), have to be independent.

In the case of the preference relation including equivalent elements, the condition (14) can be relaxed to the form (15) – (16).

The assumptions A2 – A3 reflect the following properties of distributions of comparisons errors: the probability of correct comparison is greater than of the incorrect one - in the case of binary comparisons (inequality (14)); zero is the median of each distribution of comparison error (inequalities (14) – (16)); zero is the mode of each distribution of comparison error (inequalities (14) – (18)); the set of all comparisons comprises the realizations of independent random variables; the expected value of any comparison error can differ from zero.

2.3 The Form of Estimators

Two forms of estimators are examined. Their properties have been proven by the author (see References) on the basis of the well-known probabilistic inequalities (Hoeffding, 1963, Chebyshev), properties of order statistics (David, 1970), and convergence of variances of relevant variables to zero.

The estimate based on the total sum of differences, denoted $\hat{\chi}_1^{(\ell)}, \dots, \hat{\chi}_{\hat{n}}^{(\ell)}$ (or $\hat{T}_v^{(\ell)}(x_i, x_j) < i, j > \in R_m$), results from the minimization problem:

$$\min_{\chi_1^{(\ell)}, \dots, \chi_r^{(\ell)} \in F_X^{(\ell)}} \left\{ \sum_{< i, j > \in R_m} \sum_{k=1}^N \left| g_{vk}^{(\ell)}(x_i, x_j) - t_v^{(\ell)}(x_i, x_j) \right| \right\} \quad (19)$$

where:

$F_X^{(\ell)}$ – the feasible set, i.e. the family of all relations $\chi_1^{(\ell)}, \dots, \chi_r^{(\ell)}$ of ℓ -th type in the set \mathbf{X} ,

$t_v^{(\ell)}(x_i, x_j)$ – the function describing any relation $\{\chi_1^{(\ell)}, \dots, \chi_r^{(\ell)}\}$ of ℓ -th type,

R_m – the set of the form $R_m = \{< i, j > \mid 1 \leq i, j \leq m; j > i\}$

(symbol $g_{vk}^{(\ell)}(x_i, x_j)$ is used for both random variables and realizations).

In the case of the preference relation and binary comparisons the following transformation is also applied:

$$\theta(g_{vk}^{(\ell)}(x_i, x_j) - t_v^{(\ell)}(x_i, x_j)) = \begin{cases} 0 & \text{if } g_{vk}^{(\ell)}(x_i, x_j) = t_v^{(\ell)}(x_i, x_j); \\ 1 & \text{if } g_{vk}^{(\ell)}(x_i, x_j) \neq t_v^{(\ell)}(x_i, x_j). \end{cases} \quad (20)$$

The criterion function with the use of the transformation (20) is simpler from the computational point of view. The properties of both approaches are similar (Klukowski, 1990b).

The estimate based on medians, denoted $\hat{\chi}_1^{(\ell)}, \dots, \hat{\chi}_r^{(\ell)}$ (or $\hat{T}_v^{(\ell)}(x_i, x_j)$), is obtained on the basis of the following minimization problem:

$$\min_{\chi_1^{(\ell)}, \dots, \chi_r^{(\ell)} \in F_X} \left\{ \sum_{< i, j > \in R_m} \left| g_v^{(\ell, me)}(x_i, x_j) - t_v^{(\ell)}(x_i, x_j) \right| \right\}, \quad (21)$$

where:

$g_v^{(\ell, me)}(x_i, x_j)$ – the sample median in the set $\{g_{v,1}^{(\ell)}(x_i, x_j), \dots, g_{v,N}^{(\ell)}(x_i, x_j)\}$.

The estimate, resulting from the criterion (19) or (20) will be denoted with symbols $\hat{\chi}_1^{(\ell)}, \dots, \hat{\chi}_r^{(\ell)}$ or (equivalently) $\hat{T}_v^{(\ell)}(x_i, x_j)$, while the estimate resulting from the criterion (21) – with symbols $\hat{\chi}_1^{(\ell)}, \dots, \hat{\chi}_r^{(\ell)}$ or $\hat{T}_v^{(\ell)}(x_i, x_j)$.

In the case of the preference relation and medians from comparisons, the same transformation can be also applied.

The number of estimates, resulting from the criterion functions (19), (21) can exceed one; the unique estimate can be determined in a random way or as a result of validation. Multiple estimates can appear also in other methods (see David 1988, Ch. 2).

The assumptions A1 – A3 allow for inference about distributions of errors of estimates. Let us discuss first the estimator based on of the criterion (19). For each relation type one can determine a finite set including all possible realizations of comparisons

$$g_{vk}^{(\ell)}(x_i, x_j), (\ell \in \{e, \tau, p\}, v \in \{b, \mu\}, k = 1, \dots, N; i, j \in R_m)$$

and the probability of each realization. The use of the criterion (19) determines: the estimate, its probability and estimation error. The error has the form: $\{\hat{T}_v^{(\ell)}(x_i, x_j) - T_v^{(\ell)}(x_i, x_j); i, j \in R_m\}$, i.e. it is a multidimensional random variable. The analysis of such error is, in fact, unrealizable and it is suggested to replace it with one-dimension error:

$$\hat{\Delta}_v^{(\ell)} = \sum_{i, j \in R_m} |\hat{T}_v^{(\ell)}(x_i, x_j) - T_v^{(\ell)}(x_i, x_j)|. \quad (22)$$

The estimate with the error $\hat{\Delta}_v^{(\ell)} = 0$ is the errorless estimate. The probability of such error can be determined in the analytic way – as a sum of probabilities of all realizations of comparisons indicating the errorless estimate. It is clear that its value (probability) depends on the number of comparisons N and the variance of comparison errors; increase of N decreases the probability of such error and decreases the variance of the estimator. The probabilities of errors different from zero can be determined in a similar way; all possible errors and their probabilities determine the distribution function of the estimation error. Determination of the probability function in the analytic manner is complicated and involves huge computational cost - even for moderate m . Therefore, simulation approach has to be used for this purpose.

Similar considerations apply for the criteria (20), (21).

3 Properties of Estimators

The analytical properties of the estimators, established by the author, are based on properties of random variables expressing differences between pairwise comparisons and the relation form (expressed by $T_v^{(\ell)}(x_i, x_j)$). It has been proved by the author that the variables corresponding to the actual relation form have different properties than the variables corresponding to any other relation. The following results have been obtained: (i) the expected values of the variables expressing differences between comparisons and the relation form (see e.g. (26) below), corresponding to actual relation form are lower than the expected values of variables corresponding to any other relation (see e.g. (30)); (ii) the variances of the variables expressing differences between comparisons and the relation form, both - actual and different than actual, divided by the number of comparisons N in the case of sum of differences, converge to zero for $N \rightarrow \infty$; (iii) the probability of the event that the variable corresponding to actual relation assumes a value lower than the variable corresponding to a relation other than actual converges to one for $N \rightarrow \infty$; the speed of convergence guarantees good efficiency of the estimates.

Properties (i) - (iii) provide the basis for construction of the estimators; these properties have been complemented with some additional features (Klukowski 1994) and a simulation experiment. An important result of the experiment consists

in the fact that efficiency of the estimator based on the sum of inconsistencies is higher than of the median estimator; the latter estimator is, though, simpler from computational point of view and more robust.

Let us illustrate these considerations by the simplest case, i.e. equivalence relation and the estimator resulting from the criterion (19). The differences between any comparison $g_{bk}^{(e)}(x_i, x_j)$ and $T_b^{(e)}(x_i, x_j)$ assume the form:

$$U_{bk}^{(e)*}(x_i, x_j) = \begin{cases} 0 & \text{if } g_{bk}^{(e)}(x_i, x_j) = T_b^{(e)}(x_i, x_j); \quad T_b^{(e)}(x_i, x_j) = 0; \\ 1 & \text{if } g_{bk}^{(e)}(x_i, x_j) \neq T_b^{(e)}(x_i, x_j); \quad T_b^{(e)}(x_i, x_j) = 0, \end{cases} \quad (23)$$

$$V_{bk}^{(e)*}(x_i, x_j) = \begin{cases} 0 & \text{if } g_{bk}^{(e)}(x_i, x_j) = T_b^{(e)}(x_i, x_j); \quad T_b^{(e)}(x_i, x_j) = 1; \\ 1 & \text{if } g_{bk}^{(e)}(x_i, x_j) \neq T_b^{(e)}(x_i, x_j); \quad T_b^{(e)}(x_i, x_j) = 1. \end{cases} \quad (24)$$

The sum of differences assumes, for any k ($1 \leq k \leq N$), the form:

$$\sum_{<i,j>\in I^{(e)*}} U_{bk}^{(e)*}(x_i, x_j) + \sum_{<i,j>\in J^{(e)*}} V_{bk}^{(e)*}(x_i, x_j), \quad (25)$$

where:

$I^{(e)*}$ – the set of pairs $\{<i, j> \mid T_b^{(e)*}(x_i, x_j) = 0\}$,

$J^{(e)*}$ – the set of pairs $\{<i, j> \mid T_b^{(e)*}(x_i, x_j) = 1\}$.

The total sum of the differences between the relation form and the comparisons is equal:

$$W_{bN}^{(e)*} = \sum_{k=1}^N \left(\sum_{<i,j>\in I^{(e)*}} U_{bk}^{(e)*}(x_i, x_j) + \sum_{<i,j>\in J^{(e)*}} V_{bk}^{(e)*}(x_i, x_j) \right). \quad (26)$$

Under the assumptions A1, A2, A3, the expected values of the variables $U_{bk}^{(e)*}(x_i, x_j)$, $V_{bk}^{(e)*}(x_i, x_j)$ satisfy the inequalities: $E(U_{bk}^{(e)*}(x_i, x_j)) \leq \delta$, $E(V_{bk}^{(e)*}(x_i, x_j)) \leq \delta$. Therefore, the expected value of the variable $W_{bN}^{(e)*}$ satisfies the inequality $E(W_{bN}^{(e)*}) \leq \frac{Nm(m-1)}{2}\delta$. Assumptions A1 – A3 allow for determining the variance $Var(W_{bN}^{(e)*})$; its value is finite and satisfies the inequality $Var(W_{bN}^{(e)*}) \leq \frac{Nm(m-1)}{2}\delta(1-\delta)$. Obviously:

$$E\left(\frac{1}{N}W_{bN}^{(e)*}\right) \leq \frac{m(m-1)}{2}\delta, \quad (27)$$

$$\lim_{N \rightarrow \infty} Var\left(\frac{1}{N}W_{bN}^{(e)*}\right) = 0. \quad (28)$$

Let us consider any relation $\tilde{\chi}_1^{(e)}, \dots, \tilde{\chi}_{\tilde{n}}^{(e)}$ different than $\chi_1^{(e)*}, \dots, \chi_n^{(e)*}$; this means that there exist pairs (x_i, x_j) , such that $\tilde{T}_b^{(e)}(x_i, x_j) \neq T_b^{(e)}(x_i, x_j)$. Define the

random variables $\tilde{U}_{bk}^{(e)}(x_i, x_j)$, $\tilde{V}_{bk}^{(e)}(x_i, x_j)$, $\tilde{W}_b^{(e)}(x_i, x_j)$ corresponding to the such values $\tilde{T}_b^{(e)}(x_i, x_j)$:

$$\begin{aligned}\tilde{U}_{bk}^{(e)}(x_i, x_j) &= \begin{cases} 0 & \text{if } g_{bk}^{(e)}(x_i, x_j) = \tilde{T}_b^{(e)}(x_i, x_j); \quad \tilde{T}_b^{(e)}(x_i, x_j) = 0; \\ 1 & \text{if } g_{bk}^{(e)}(x_i, x_j) \neq \tilde{T}_b^{(e)}(x_i, x_j); \quad \tilde{T}_b^{(e)}(x_i, x_j) = 0, \end{cases} \\ \tilde{V}_{bk}^{(e)}(x_i, x_j) &= \begin{cases} 0 & \text{if } g_{bk}^{(e)}(x_i, x_j) = \tilde{T}_b^{(e)}(x_i, x_j); \quad \tilde{T}_b^{(e)}(x_i, x_j) = 1; \\ 1 & \text{if } g_{bk}^{(e)}(x_i, x_j) \neq \tilde{T}_b^{(e)}(x_i, x_j); \quad \tilde{T}_b^{(e)}(x_i, x_j) = 1, \end{cases} \quad (29)\end{aligned}$$

$$\tilde{W}_{bN}^{(e)} = \sum_{k=1}^N \left(\sum_{\tilde{I}^{(e)}} \tilde{U}_{bk}^{(e)}(x_i, x_j) + \sum_{\tilde{J}^{(e)}} \tilde{V}_{bk}^{(e)}(x_i, x_j) \right). \quad (30)$$

where:

$\tilde{I}^{(e)}$ – the set of pairs $\{<i, j> \mid \tilde{T}_b^{(e)}(x_i, x_j) = 0\}$,

$\tilde{J}^{(e)}$ – the set of pairs $\{<i, j> \mid \tilde{T}_b^{(e)}(x_i, x_j) = 1\}$.

The expected values $E(\tilde{U}_{bk}^{(e)}(x_i, x_j))$, $E(\tilde{V}_{bk}^{(e)}(x_i, x_j))$ assume the form:

$$\begin{aligned}E(\tilde{U}_{bk}^{(e)}(x_i, x_j)) &= 0 * P(g_{bk}^{(e)}(x_i, x_j) = 0 \mid T_b^{(e)}(x_i, x_j) = 1) + \\ &\quad 1 * P(g_{bk}^{(e)}(x_i, x_j) = 1 \mid T_b^{(e)}(x_i, x_j) = 1) \geqslant 1 - \delta, \quad (31)\end{aligned}$$

$$\begin{aligned}E(\tilde{V}_{bk}^{(e)}(x_i, x_j)) &= 0 * P(g_{bk}^{(e)}(x_i, x_j) = 0 \mid T_b^{(e)}(x_i, x_j) = 0) + \\ &\quad 1 * P(g_{bk}^{(e)}(x_i, x_j) = 1 \mid T_b^{(e)}(x_i, x_j) = 0) \geqslant 1 - \delta, \quad (32)\end{aligned}$$

and:

$$E(\tilde{W}_{bN}^{(e)}) = \sum_{k=1}^N \left(\sum_{\tilde{I}^{(e)}} (\tilde{U}_{bk}^{(e)}(x_i, x_j)) + \sum_{\tilde{J}^{(e)}} (\tilde{V}_{bk}^{(e)}(x_i, x_j)) \right) > \frac{m(m-1)}{2} \delta. \quad (33)$$

The formulae (27) – (33) indicate that the expected value $E(\frac{1}{N} W_{bN}^{(e)*})$, corresponding to the actual relation $\chi_1^{(e)*}, \dots, \chi_n^{(e)*}$, is lower than the expected value $E(\frac{1}{N} \tilde{W}_{bN}^{(e)})$, corresponding to any other relation $\tilde{\chi}_1^{(e)}, \dots, \tilde{\chi}_{\tilde{n}}^{(e)}$. The variances of both variables converge to zero for $N \rightarrow \infty$. The variables $U_{bk}^{(e)*}(x_i, x_j)$, $V_{bk}^{(e)*}(x_i, x_j)$ assume values equal to $|g_{bk}^{(e)}(x_i, x_j) - T_b^{(e)}(x_i, x_j)|$, used in the criterion function (19). Moreover, it can be also shown (see Klukowski, 1994), that:

$$P(W_{bN}^{(e)*} < \tilde{W}_{bN}^{(e)}) \geqslant 1 - \exp\left\{-2N\left(\frac{1}{2} - \delta\right)^2\right\}. \quad (34)$$

The above facts indicate that the estimator $\hat{\chi}_1^{(e)}, \dots, \hat{\chi}_{\hat{n}}^{(e)}$, minimizing the number of inconsistencies with comparisons, guarantees the errorless estimate for $N \rightarrow \infty$. The inequality (34) shows that the errorless estimate can be obtained with the probability close to one for finite N and indicates the influence of δ and N on the precision of the estimator.

The properties of the median estimator are similar, especially (see Klukowski, 1994):

$$P(W_{bN}^{(p,me)*} < \tilde{W}_{bN}^{(me,p)}) \geq 1 - 2\exp\{-2N(\frac{1}{2} - \delta)^2\}.$$

The case of multivalent comparisons, can be analyzed in a similar way the details are presented in Klukowski 2011a, Chap. 6 and 8.

4 Solving of Optimization Problems

Minimization of the functions (19), (21) is, in general, not an easy problem, because of the dimensions of the feasible set. Currently, the algorithms are available only for ranking problems based on binary single comparisons (see David, 1988, Chapt. 2, Hansen P., et al 1994); they refer to the dynamic programming or branch-and-bound algorithms, some of them can be used for known n . The algorithms are efficient for the moderate number of elements m . In the case of large m , the problems can be also solved with the use of heuristic algorithms: genetic (Falkenauer, 1998), artificial neural networks, random search (Ripley, 2006), etc.

In the case of multivalent comparisons the exact algorithms are not available now. The problems with moderate number of elements m , i.e. 3 – 12, can be solved with the use of complete enumeration. Problems with higher number of elements can be solved using heuristic algorithms, mentioned above.

It is obvious that the estimators based on multivalent comparisons require more computations than those based on binary comparisons.

5 Summary – Achievements of the Work and Further Researches

The work presents the synthesis of main results, of the author (Klukowski 2011a), concerning estimation of three relations – equivalence, tolerance, and preference – on the basis of pairwise comparisons with random errors (see Klukowski in References). The problems of that type occur often in applications and have been investigated in literature.

The following new results, presented here, should be emphasized.

1⁰. Two types of data have been taken into account: binary and multivalent.

2⁰. The assumptions concerning the comparison errors are weaker than those commonly used in the literature.

3⁰. Two estimators have been examined; the first one is based on the sum of differences between the relation form and the comparison data, the second is based on differences between the relation form and the median from comparisons of each pair. The estimators have analytical properties guaranteeing good efficiency.

4⁰. The analytical properties of the estimators have been complemented with the results of simulation study (Klukowski 2011a). This allows for determining of

parameters, especially the number of comparisons N , guaranteeing the required precision of estimates; a definite value of N provides for the frequency of errorless result close to one or equal one.

5⁰. The properties of estimates can be thoroughly validated (Klukowski 2011a, Chapt. 10); validation comprises the fact of existence of the relation and the assumptions as to the comparison errors.

6⁰. The approach proposed allows for combining of comparisons obtained from different sources, e.g. statistical tests, experts, neural networks. It is also possible to combine binary and multivalent data and to apply two-stage estimators, based, in the first stage, on binary comparisons, and in the second stage – on multivalent comparisons, obtained in the first stage.

7⁰. The approach presented will be developed in the following directions: statistical learning, estimation of more complex structures of data, e.g. hierarchical, multiple criteria comparisons, etc. An important field is also constituted by application of the estimators and tests developed.

References

1. Bradley, R.A.: Science, statistics and paired comparisons. *Biometrics* 32, 213–232 (1976)
2. Bradley, R.A.: Paired comparisons: some basic procedures and examples. In: Krishnaiyah, P.R., Se, P.K. (eds.) *Handbook of Statistics*, vol. 4, pp. 299–326. North-Holland, Amsterdam (1984)
3. Brunk, H.D.: Mathematical models for ranking from paired comparisons. *JASA* 55, 503–520 (1960)
4. David, H.A.: *Order Statistics*. J.Wiley (1970)
5. David, H.A.: *The Method of Paired Comparisons*, 2nd edn. Ch. Griffin, London (1988)
6. Davidson, R.R.: A bibliography on the Method of Paired Comparisons. *Biometrics* 32, 241–251 (1976)
7. Falkenauer, E.: *Genetics Algorithms and Grouping Data*. J. Wiley (1998)
8. Gordon, A.D.: *Classification*, 2nd edn. Chapman&Hall/CRC (1999)
9. Hand, D.J.: *Discrimination and Classification*. J. Wiley (1986)
10. Hartigan, J.J.: *Clustering Algorithms*. J. Wiley (1975)
11. Hansen, P., Jaumard, B., Sanlaville, E.: Partitioning Problems in Cluster Analysis: A Review of Mathematical Programming Approaches. In: *Studies in Classification, Data Analysis, And Knowledge Organization*. Springer (1994)
12. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. Springer (2002)
13. Hoeffding, W.: Probability inequalities for sums of bounded random variables. *JASA* 58, 13–30 (1963)
14. Kaufman, L., Rousseeuw, P.J.: *Findings Groups In Data: An Introduction to Cluster Analysis*. J. Wiley (1990)
15. Klukowski, L.: Algorithm for classification of samples in the case of unknown number variables generating them. *Przeglad Statystyczny XXXVII*, 167–177 (1990a) (in Polish)
16. Klukowski, L.: Ranking of items on the basis of pairwise comparisons in the case of random errors. In: Kulikowski, R., Kacprzyk, J. (eds.) *SYSTEMS RESEARCH*, vol. 3, pp. 212–268. Omnitech Press, Warsaw (1990b) (in Polish)

17. Klukowski, L.: Some probabilistic properties of the nearest adjoining order method and its extensions. *Annals of Operational Research* 51, 241–261 (1994)
18. Klukowski, L.: The nearest adjoining order method for pairwise comparisons in the form of difference of ranks. *Annals of Operations Research* 97, 357–378 (2000)
19. Klukowski, L.: Estimation of the Preference Relation on the Basis of Medians from Pairwise Comparisons in the Form of Difference of Ranks. In: Kurzynski, et al. (eds.) *Computer Recognition Systems 2. AISC*, vol. 45, pp. 232–241. Springer, Heidelberg (2007a)
20. Klukowski, L.: Estimation of tolerance relation the basis of multiple pairwise comparisons with random errors. *Control and Cybernetics* 36, 443–466 (2007b)
21. Klukowski, L.: Estimation of the preference relation the basis of multiple pairwise comparisons in the form of differences of ranks. *Control and Cybernetics* 37, 711–729 (2008)
22. Klukowski, L.: Methods of Estimation of Relations of: Equivalence, Tolerance, and Preference in a Finite Set. Warsaw. IBS PAN, Series: Systems Research, vol. 69 (2011a)
23. Klukowski, L.: Estimation of Tolerance Relation on the Basis of Pairwise Comparisons. In: Burduk, R., et al. (eds.) *Computer Recognition Systems 4. AISC*, vol. 95, pp. 232–241. Springer, Heidelberg (2011b)
24. Kohonen, T.: *Self-Organizing Maps*. Springer (1995)
25. Ripley, B.D.: *Stochastic Simulation*. J. Wiley (2006)
26. Slater, P.: Inconsistencies in a schedule of paired comparisons. *Biometrika* 48, 303–312 (1961)

Generalized Constraint Design of Linear-Phase FIR Digital Filters

Norbert Henzel and Jacek M. Leski

Abstract. We consider the design of digital finite-impulse response (FIR) filters satisfying constraints on the amplitude response. Constrained FIR filter design with frequency-domain linear constraints on the amplitude response usually uses the least-squares or the Chebyshev error criterion, which can be generally reformulated as quadratic programming (QP) problem. This paper presents an novel algorithm for the design of constrained low-pass FIR filters according to a variously defined error. This approach does not require the transition bands specification, is characterized by rapid convergence and is suitable for high order filter design.

Keywords: Biomedical Signal Processing, Digital Filter Design, FIR Filters.

1 Introduction

Linear-phase finite impulse response (FIR) digital filters play a crucial role in a large number of signal processing problems, for example, biomedical signal processing, image processing, telecommunication application, etc. Therefore, linear-phase filters design methods have been widely explored [1].

The FIR filter design, in majority of cases, can be regarded as an optimization problem, where a desired, ideal frequency response is approximated.

The FIR filter design process typically require several steps. First step consists in defining a desired, ideal, frequency response. Second step demands selecting the

Norbert Henzel

Institute of Electronics, Silesian University of Technology, 44-101 Gliwice,
Akademicka 16, Poland, Institute of Medical Technology and Equipment,
41-800 Zabrze, Roosevelta 118, Poland
e-mail: nhenzel@polsl.pl

Jacek M. Leski

Institute of Electronics, Silesian University of Technology, 44-101 Gliwice,
Akademicka 16, Poland
e-mail: jleski@polsl.pl

desired filter length. Third, establishing a measure of error between desired and obtained filter frequency responses (different optimality criteria result in different filter behavior). Fourth step consist in applying (or developing) an optimization method to find the filter coefficients. In scientific literature a vast number of different techniques for designing digital FIR filters have been proposed and the majority of them uses one (or a combination) of the following error criteria [2], [4]: least-squares ([5], [15], [17]), Chebyshev (minimax) ([10], [12]) and maximally flat ([16]).

The constrained design of FIR filters using the least squares error function for the first time was presented in [2]. This approach to FIR filter design has also been next considered and further developed in a vast number of papers, e.g. [3], [6], [13], [14].

The goal of this paper is to present a new method of constrained FIR filter design and to investigate its performances for different design problems. Although presented in the context of low-pass filter design it can be easily extended to other FIR filter design problems.

2 Preliminaries

A digital filter is a linear time-invariant system, operating on an input sequence $x(n)$ to produce an output sequence $y(n)$, where n denotes discrete time. This system can be completely described by the impulse response sequence $h(n)$. The input-output relation for digital filter is given by [11], [8]

$$y(k) = \sum_{m=-\infty}^{\infty} x(m)h(k-m) = \sum_{m=-\infty}^{\infty} x(k-m)h(m). \quad (1)$$

Typically, $h(m) = 0$ for $0 > m > N - 1$, so we obtain

$$y(k) = \sum_{m=0}^{N-1} x(k-m)h(m). \quad (2)$$

The number of impulse response coefficients, N , is said to be the length of the filter, and the quantity $N - 1$ is called the order of the filter [1].

The frequency response $H(e^{j\omega})$ of an FIR filter is given by the discrete-time Fourier transform of its impulse response $h(n)$ [1]:

$$H(e^{j\omega}) = \sum_{n=0}^{N-1} h(n) e^{-j\omega n} \quad (3)$$

where the frequency $\omega \in [0, \pi]$.

If the impulse response $h(n)$ of the FIR filter has even symmetry, $h(n) = h(N - 1 - n)$, or odd symmetry, $h(n) = -h(N - 1 - n)$, the phase response of the designed filter is linear and the obtained design problem is real-valued. In this case, the frequency response function $H(e^{j\omega})$ can be written as [1]

$$H(e^{j\omega}) = e^{-j(N-1)/2\omega} e^{-j\beta} H_0(\omega) \quad (4)$$

where $H_0(\omega)$ is a real-valued function, called amplitude response and the constant β satisfies $\beta = 0$ or $\beta = \pi/2$. In the first case, $\beta = 0$, the filter amplitude response is given by [1]

$$H_0(\omega) = \begin{cases} \sum_{n=0}^{(N-1)/2} b_n \cos(\omega n) & \text{for } N-1 \text{ even,} \\ \sum_{n=0}^{N/2} b_n \cos(\omega(n - \frac{1}{2})) & \text{for } N-1 \text{ odd.} \end{cases} \quad (5)$$

where the coefficients b_n are related to $h(n)$ in as follows:

$$b_n = \begin{cases} h\left(\frac{N-1}{2}\right) & \text{for } N-1 \text{ even, } n=0, \\ 2h\left(\frac{N-1}{2} - n\right) & \text{for } N-1 \text{ even, } n \neq 0, \\ 2h\left(\frac{N}{2} - n\right) & \text{for } N-1 \text{ odd.} \end{cases} \quad (6)$$

Similar expressions can be developed for $\beta = \pi/2$ [1].

The linear phase response of a FIR filter is a very desirable property in many applications, e.g. processing of an electrocardiographic (ECG) signals, acoustics signals, etc.

Low-pass FIR digital filters are characterized by: the length of the impulse response N , the passband edge frequency f_p , the stopband edge frequency f_s , the maximum passband ripple (maximum passband gain) δ_p and minimum stopband attenuation (maximum stopband gain) δ_s . The last two values are often expressed in decibels:

$$D_p = 20 \log_{10}(\delta_p) \text{ [dB]}, \quad D_s = 20 \log_{10}(\delta_s) \text{ [dB]}. \quad (7)$$

Figure 1 depicts the filter parameters and performance measures discussed so far.

The relation between the linear-phase FIR filter amplitude response $H_0(\omega)$ for $\beta = 0$ and $N-1$ even, and the coefficients b_n for a given set of frequency points ω_i , $i = 1, \dots, L$, distributed over the frequency domain can be compactly represented in matrix form. For example, the first case in (5) can be written as

$$\mathbf{H}_0 \triangleq [H_0(\omega_1), H_0(\omega_2), \dots, H_0(\omega_L)]^\top = \mathbf{T}\mathbf{b}, \quad (8)$$

where

$$\mathbf{b} = [b_0, b_1, \dots, b_M]^\top; \quad M = (N-1)/2, \quad (9)$$

$$H_0(\omega_i) = \mathbf{b}^\top \mathbf{t}(\omega_i), \quad (10)$$

$$\mathbf{t}(\omega_i) = [\cos(0\omega_i), \cos(1\omega_i), \dots, \cos(M\omega_i)]^\top, \quad (11)$$

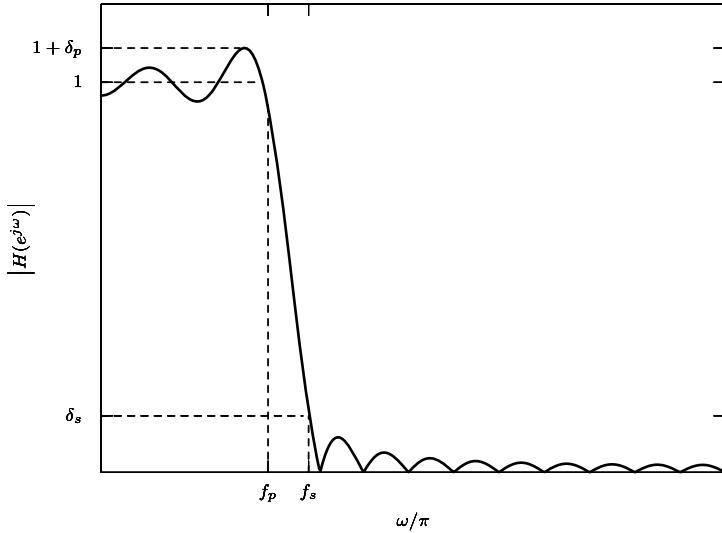


Fig. 1 Filter parameters and performance measures

$$J(\mathbf{b}) = \sum_{i=1}^L \left[\mathbf{b}^\top \mathbf{t}(\omega_i) - H_D(\omega_i) \right]^2 \quad (12)$$

and

$$\mathbf{T} = \begin{bmatrix} \cos(0 \cdot \omega_1) & \cos(\omega_1) & \cdots & \cos(M \cdot \omega_1) \\ \cos(0 \cdot \omega_2) & \cos(\omega_2) & \cdots & \cos(M \cdot \omega_2) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(0 \cdot \omega_L) & \cos(\omega_L) & \cdots & \cos(M \cdot \omega_L) \end{bmatrix} = \begin{bmatrix} \mathbf{t}(\omega_1)^\top \\ \mathbf{t}(\omega_2)^\top \\ \vdots \\ \mathbf{t}(\omega_L)^\top \end{bmatrix}. \quad (13)$$

Lets us define an error as

$$E(\omega_i) = H_0(\omega_i) - H_D(\omega_i) \quad (14)$$

where $H_0(\omega_i)$ and $H_D(\omega_i)$, $i = 1, \dots, L$, are the actual and the desired frequency response of the filter, respectively.

Now, the weighted square error criterion can be written as:

$$J(\mathbf{b}) = \|\mathbf{E}\|_2^2 = \mathbf{E}^\top \mathbf{E} = (\mathbf{T}\mathbf{b} - \mathbf{H}_D)^\top (\mathbf{T}\mathbf{b} - \mathbf{H}_D), \quad (15)$$

where

$$\mathbf{H}_D = [H_D(\omega_1), H_D(\omega_2), \dots, H_D(\omega_L)]^\top \quad (16)$$

is a real-valued desired amplitude response vector.

3 New Method for Low-Pass FIR Filter Design

In the proposed method of FIR filter design various loss functions are used. We seek vector \mathbf{b} by the following minimization

$$\min_{\mathbf{b} \in \mathbb{R}^{M+1}} J(\mathbf{b}) \triangleq \sum_{i=1}^L g_i \mathcal{L} \left(\mathbf{b}^\top \mathbf{t}(\omega_i) - H_D(\omega_i) \right), \quad (17)$$

where $\mathcal{L}(\cdot)$ stands for a loss function used, and g_i is a weight corresponding to the i th frequency. If we choose the quadratic loss function then in matrix notation (17) takes the form

$$\min_{\mathbf{b} \in \mathbb{R}^{M+1}} J(\mathbf{b}) \triangleq (\mathbf{T}\mathbf{b} - \mathbf{H}_D)^\top \mathbf{G} (\mathbf{T}\mathbf{b} - \mathbf{H}_D), \quad (18)$$

where $\mathbf{G} = \text{diag}(g_1, g_2, \dots, g_L)$. The role of g_i s parameters may be twofold: (i) they may correspond to our weight of the i th frequency (${}^c g_i \in [0, 1]$), (ii) through the proper selection of the parameters values we may change various error functions to the quadratic loss (${}^l g_i \in \mathbb{R}^+ \cup \{0\}$). In the last case, the values of the parameters depend on the obtained residuals. In turn, the residuals depend on \mathbf{b} . Thus, criterion function (18) should only be minimized by iteratively reweighting scenario. Let us denote \mathbf{b} , \mathbf{G} and \mathbf{E} in the k th iteration as $\mathbf{b}^{(k)}$, $\mathbf{G}^{(k)}$ and $\mathbf{E}^{(k)}$, respectively. Criterion function (18) for the k th iteration takes the form

$$J^{(k)} \left(\mathbf{b}^{(k)} \right) \triangleq \left(\mathbf{T}\mathbf{b}^{(k)} - \mathbf{H}_D \right)^\top \mathbf{G}^{(k)} \left(\mathbf{T}\mathbf{b}^{(k)} - \mathbf{H}_D \right), \quad (19)$$

where the elements on the main diagonal of $\mathbf{G}^{(k)} = \text{diag} \left(g_1^{(k)}, g_2^{(k)}, \dots, g_N^{(k)} \right)$ depend on the residuals from the previous iteration

$$\mathbf{E}^{(k-1)} = \mathbf{T}\mathbf{b}^{(k-1)} - \mathbf{H}_D. \quad (20)$$

and take the form

$$g_i^{(k)} = {}^c g_i \cdot {}^l g_i^{(k)}. \quad (21)$$

Parameter ${}^c g_i$, representing *a priori* confidence to the i th frequency does not depend on the iteration index k . In contrast, parameter ${}^l g_i^{(k)}$ depends on the i th residual from the previous iteration, $(k-1)$ th. The following form of ${}^l g_i^{(k)}$ is proposed

$${}^l g_i^{(k)} = \begin{cases} 0, & E(\omega_i)^{(k-1)} = 0, \\ \mathcal{L} \left(E(\omega_i)^{(k-1)} \right) / \left(E(\omega_i)^{(k-1)} \right)^2, & E(\omega_i)^{(k-1)} \neq 0. \end{cases} \quad (22)$$

Indeed, for the quadratic loss function, we obtain ${}^l g_i^{(k)} = 1$, for all $i = 1, 2, \dots, L$; $k = 1, 2, 3, \dots$. The absolute error function is easy to obtain by taking [7]

$${}^l g_i^{(k)} = \begin{cases} 0, & E(\omega_i)^{(k-1)} = 0, \\ 1 / |E(\omega_i)^{(k-1)}|, & E(\omega_i)^{(k-1)} \neq 0. \end{cases} \quad (23)$$

Many other loss functions easily may be obtained (see [7]).

To start this sequential optimization, we set the weights in the 0th iteration as ${}^l g_i^{(0)} = 1$ for all i .

The ε -insensitive loss function disregards errors below some $\varepsilon > 0$, chosen a priori:

$$\mathcal{L}(\zeta) = \begin{cases} 0, & |\zeta| \leq \varepsilon, \\ |\zeta| - \varepsilon, & |\zeta| > \varepsilon. \end{cases} \quad (24)$$

Various ε -insensitive loss functions may be considered, including ε -insensitive quadratic, ε -insensitive Huber, and so on. Let us start our consideration from the ε -insensitive quadratic loss

$$\mathcal{L}(\zeta) = \begin{cases} 0, & |\zeta| - \varepsilon \leq 0, \\ (\varepsilon - \zeta)^2, & \varepsilon - \zeta < 0, \\ (\varepsilon + \zeta)^2, & \varepsilon + \zeta < 0. \end{cases} \quad (25)$$

Taking into account the above equation (17), assuming $g_i = 1$ for all $i = 1, 2, \dots, L$, may be written as

$$\sum_{i=1}^L \mathcal{L}\left(\mathbf{b}^\top \mathbf{t}(\omega_i) - H_D(\omega_i)\right) = \sum_{i=1}^L g_i^+ \left(-\mathbf{b}^\top \mathbf{t}(\omega_i) + H_D(\omega_i) + \varepsilon^+(\omega_i)\right)^2 + \sum_{i=1}^L g_i^- \left(\mathbf{b}^\top \mathbf{t}(\omega_i) - H_D(\omega_i) + \varepsilon^-(\omega_i)\right)^2, \quad (26)$$

where g_i^+ (g_i^-) are equal to zero for $-\mathbf{b}^\top \mathbf{t}(\omega_i) + H_D(\omega_i) + \varepsilon^+(\omega_i) \geq 0$ ($\mathbf{b}^\top \mathbf{t}(\omega_i) - H_D(\omega_i) + \varepsilon^-(\omega_i) \geq 0$) and 1 otherwise. Thus, the ε -insensitive quadratic loss function may be decomposed into two asymmetric quadratic loss functions. Let \mathbf{T}_e be the $2L \times (M+1)$ matrix

$$\mathbf{T}_e^\top \triangleq \left[\mathbf{T}^\top, -\mathbf{T}^\top \right] \quad (27)$$

and \mathbf{H}_{De} be the $2L$ -dimensional vector $\mathbf{H}_{De}^\top = [\mathbf{H}_D^\top - \boldsymbol{\varepsilon}^+, -\mathbf{H}_D^\top - \boldsymbol{\varepsilon}^-]$ and $\boldsymbol{\varepsilon}^+ = [\varepsilon^+(\omega_1), \varepsilon^+(\omega_2), \dots, \varepsilon^+(\omega_L)]^\top$, $\boldsymbol{\varepsilon}^- = [\varepsilon^-(\omega_1), \varepsilon^-(\omega_2), \dots, \varepsilon^-(\omega_L)]^\top$. Using the above mentioned notation, criterion function (19) for k th iteration takes the form

$$J^{(k)}\left(\mathbf{b}^{(k)}\right) \triangleq \left(\mathbf{T}_e \mathbf{b}^{(k)} - \mathbf{H}_{De}\right)^\top \mathbf{G}^{(k)} \left(\mathbf{T}_e \mathbf{b}^{(k)} - \mathbf{H}_{De}\right), \quad (28)$$

where the elements on the main diagonal of $\mathbf{G}^{(k)}$ (now, $(2L) \times (2L)$ matrix) depend on residuals from the previous iteration

$$\mathbf{E}^{(k-1)} = \mathbf{T}_e \mathbf{b}^{(k-1)} - \mathbf{H}_{De}. \quad (29)$$

The fitting of the i th frequency is represented by the i th and the $(i+L)$ th element of \mathbf{E} . If both $E(\omega_i)^{(k)}$ and $E(\omega_{i+L})^{(k)}$ are greater than or equal to zero, then characteristics for the i th frequency falls, for the k th iteration, into the insensibility zone, i.e., $\mathbf{H}_D - \boldsymbol{\epsilon}^- \preceq \mathbf{H}_0 \preceq \mathbf{H}_D + \boldsymbol{\epsilon}^+$, where the symbol \preceq stands for componentwise inequality. If $E(\omega_i)^{(k)}$ ($E(\omega_{i+L})^{(k)}$) is less than zero, then the i th frequency is below (above) the insensibility zone in the k th iteration and should be penalized. For the ε -insensitive quadratic (ε SQ) loss we have

$${}^l g_i^{(k)} = \begin{cases} 0, & E(\omega_i)^{(k-1)} \geq 0, \\ 1, & E(\omega_i)^{(k-1)} < 0. \end{cases} \quad (30)$$

Other ε -insensitive loss functions easily may be obtained, for example [7]:

- HUBer (ε HUB) with parameter $\delta > 0$

$${}^l g_i^{(k)} = \begin{cases} 0, & e_i^{(k-1)} \geq 0, \\ 1/\delta^2, & 0 > e_i^{(k-1)} \geq -\delta, \\ -1/\left(\delta |e_i^{(k-1)}|\right), & e_i^{(k-1)} < -\delta. \end{cases} \quad (31)$$

- SIGmoidal (ε SIG) with parameters $\alpha, \beta > 0$

$${}^l g_i^{(k)} = \begin{cases} 0, & e_i^{(k-1)} \geq 0, \\ 1/\left(\left(e_i^{(k-1)}\right)^2 \left(1 + \exp\left(\alpha\left(e_i^{(k-1)} + \beta\right)\right)\right)\right), & e_i^{(k-1)} < 0. \end{cases} \quad (32)$$

Our *a priori* confidence to the i th datum (${}^c g_i \in [0, 1]$) should be 'doubled', i.e., ${}^c g_{i+L} = {}^c g_i$, for $i = 1, 2, \dots, L$, because every frequency in criterion function (28) is also doubled.

The optimality condition for the k th iteration is obtained by differentiating (19) with respect to \mathbf{b} and setting the result equals to zero

$$\left(\mathbf{T}_e^\top \mathbf{G}^{(k)} \mathbf{T}_e\right) \mathbf{b}^{(k)} = \mathbf{T}_e^\top \mathbf{G}^{(k)} \mathbf{H}_{De}. \quad (33)$$

The procedure of Iteratively Reweighted least square error minimization for Constrained Filter Design (IRCFD) can be summarized in the following steps [7]:

4 Design Examples

The proposed method has been used to solve many filter design problems. The obtained results show that the proposed method is competitive when compared with results obtained by other methods. In this section, we only present the results obtained for three examples. The computation was performed in MatLab® computing environment. Two other methods were used to compare the results of the proposed method: the method proposed by McClellan, Parks and Rabiner in [9] (MPR) and

the method presented by Selesnick, Lang and Burrus in [13] (SLB). It should be noted that the MPR method is not a constraint design method and the resulting passband ripples and stopband attenuation depends on the inclination of the transition band (part of the filter frequency characteristics between f_p and f_s , see Figure 1).

-
- 1: Fix $\tau > 0$ and $\mathbf{G}^{(0)} = \mathbf{I}$. Set the iteration index $k = 0$.
 - 2: Obtain $\mathbf{b}^{(k)}$ by using conjugate gradient method to (33).
 - 3: $\mathbf{E}^{(k)} = \mathbf{T}_e \mathbf{b}^{(k)} - \mathbf{H}_{De}$.
 - 4: $\mathbf{G}^{(k+1)} = \text{diag} \left(g_1^{(k+1)}, g_2^{(k+1)}, \dots, g_N^{(k+1)} \right)$, where each $g_i^{(k+1)}$, for $i = 1, 2, \dots, L$ is obtained by (21) and depends on the selected loss function.
 - 5: **if** $k > 1$ and $\left\| \mathbf{b}^{(k)} - \mathbf{b}^{(k-1)} \right\|_2 < 10^{-3}$ **then**
 - 6: stop
 - 7: **else**
 - 8: $k \leftarrow k + 1$
 - 9: goto 2.
 - 10: **end if**
-

Example 1: We first consider the low-pass filter described in [13]: $\omega_p = 0.2728\pi$, $\omega_s = 0.3270\pi$, $\delta_p = \delta_s = 0.02$ and $N = 61$. Figure 2(a) shows the frequency responses of the obtained filter using the ε SQ error function. A comparison with the two selected methods is given in Table 1(a).

Table 1 Parameters of the frequency response for *Example 1* (a) and *Example 2* (b) low-pass filters

	(a)			(b)		
	MPR	SLB	IRCFD(ε SQ)	MPR	SLB	IRCFD(ε SQ)
E_p	0.0198	0.0039	0.0041	0.1142	0.1079	0.0136
δ_p^{max}	0.0168	0.0143	0.0134	0.0770	0.0574	0.0287
D_s^{min} [dB]	-35.4	-28.0	-28.3	-22.2	-15.7	-14.5

The proposed method has the smallest maximum ripple (δ_p^{max}) and the second smallest error in the pass band (E_p). The minimum attenuation in stopband (D_s^{min}) is comparable with the error of the SLB method; the best D_s^{min} is obtained with the MPR method.

Example 2: Next we consider the narrow low-pass filter described in [2]: $\omega_p = 0.0625\pi$, $\omega_s = 0.0804\pi$, $\delta_p = 0.0575$, $\delta_s = 0.006$ dB and $N = 95$. Figure 2(b) shows the frequency responses of the obtained filter using the ε SQ error function. A comparison with the two selected methods is given in Table 1(b).

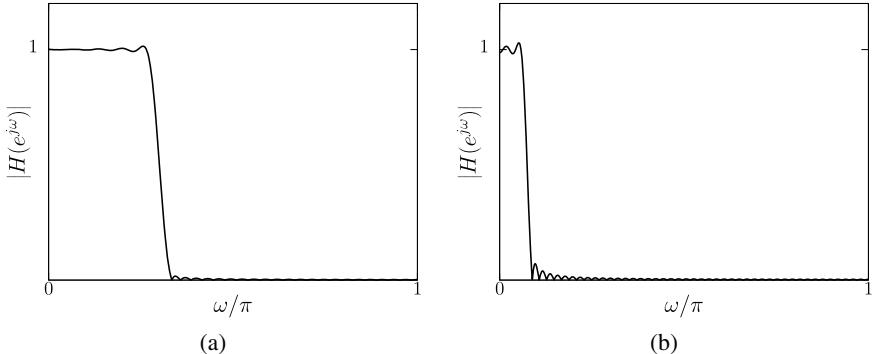


Fig. 2 Frequency responses for *Example 1* (a) and *Example 2* (b) low-pass filters

In this case the proposed method has the smallest maximum ripple (δ_p^{\max}) and the smallest error in the passband (E_p). The minimum attenuation in stopband (D_s^{\min}) is comparable with this of the SLB method; the best D_s^{\min} is given once again by the MPR method.

Example 3: In the last example we consider another narrow low-pass filter described in [2]: $\omega_p = 0.021\pi$, $\omega_s = 0.03125\pi$, $\delta_p = 0.0575$, $\delta_s = 0.0334$ and $N = 128$. A comparison with the two selected methods is given in Table 2.

Table 2 Parameters of the frequency response for *Example 3* low-pass filter

	MPR	SLB	IRCFD(ε SIG)	IRCFD(ε HUB)
E_p	0.0751	0.0238	0.0091	0.0001
δ_p^{\max}	0.1172	0.0421	0.0460	0.0074
D_s^{\min} [dB]	-18.6	-12.7	-7.8	-3.8

The smallest maximum ripple (δ_p^{\max}) and the smallest error in the passband (E_p) are obtained with the IRCFD(ε HUB) method at the expense of (D_s^{\min}). The (δ_p^{\max}) of SLB and IRCFD(ε SIG) methods are comparable but the E_p of IRCFD(ε SIG) is much better. The best D_s^{\min} is obtained for the MPR method. Once again it is observed that the smaller E_p or (δ_p^{\max}), the smaller D_s^{\min} .

5 Conclusion

In this paper we developed a FIR filter design method based on a discrete desired filter frequency response and the procedure of Iteratively Reweighted least square

error minimization for Constrained Filter Design (IRCFD). From our included examples it could be concluded that the proposed method is efficient and competitive with respect to other well-known from scientific literature methods. Additionally, the proposed method offers the possibility to select different error measures, appropriate for the design problem.

References

1. Parks, T.W., Burrus, C.S.: Digital Filter Design. Wiley, New York (1987)
2. Adams, J.W.: FIR Digital Filters with Least-Squares Stopbands Subject to Peak-Gain Constraints. *IEEE Transactions on Circuits and Systems* 39, 376–388 (1991)
3. Adams, J.W., Sullivan, J.L.: Peak-constrained least squares optimization. *IEEE Transactions on Signal Processing* 46, 306–321 (1998)
4. Burrus, C.S., Barreto, A.: Least p-power error design of filters. In: Proc. IEEE Int. Symp. Circuits. Syst. (ISCAS), vol. 1, pp. 545–548 (1992)
5. Burrus, C.S., Soewito, A.W., Gopinath, R.A.: Least squared error FIR filter design with transition bands. *IEEE Trans. Signal Process.* 40, 1327–1340 (1992)
6. Lang, M., Selesnick, I.W., Burrus, C.S.: Constrained LS design of 2D FIR filters. *IEEE Transactions on Signal Processing* 44, 1234–1241 (1996)
7. Leski, J., Henzel, N.: Generalized Ordered Linear Regression with Regularization. *Bull. Pol. Ac.: Tech.* (in print)
8. McClellan, J., Parks, T.: A unified approach to the design of optimum FIR linear-phase digital filters. *IEEE Transactions on Circuits and Systems* 20, 679–701 (1973)
9. McClellan, J., Parks, T., Rabiner, L.: A computer program for designing optimum FIR linear phase digital filters. *IEEE Transactions on Audio and Electroacoustics* 21, 506–526 (1973)
10. McClellan, J., Parks, T.: A unified approach to the design of optimum FIR linear-phase digital filter. *IEEE Trans. Circuit Theory* 20, 697–701 (1973)
11. Parks, T., McClellan, J.: A program for the design of linear phase finite impulse response digital filters. *IEEE Transactions on Audio and Electroacoustics* 20, 195–199 (1972)
12. Rabiner, L.R., McClellan, J.H., Parks, T.W.: FIR digital filter design techniques using weighted Chebyshev approximations. *Proc. IEEE* 63, 595–610 (1975)
13. Selesnick, I.W., Lang, M., Burrus, C.S.: Constrained least square design of FIR filters without specified transition bands. *IEEE Transactions on Signal Processing* 44, 1879–1892 (1996)
14. Selesnick, I.W., Lang, M., Burrus, C.S.: A modified algorithm for constrained least square design of multiband FIR filters without specified transition bands. *IEEE Transactions on Signal Processing* 46, 497–501 (1998)
15. Tufts, D.W., Francis, J.T.: Designing digital low-pass filters- comparison of some methods and criteria. *IEEE Trans. Audio Electroacoust.* 18, 487–494
16. Vaidyanathan, P.P.: Efficient and multipliers design of FIR filters with very sharp cutoff via maximally flat building blocks. *IEEE Trans. Circuits Syst.* 32, 236–244
17. Vaidyanathan, P.P., Nguyen, T.Q.: Eigenfilters: A new approach to least squares FIR filter design and applications including Nyquist filters. *IEEE Trans. Circuits Syst.* 34, 11–23 (1987)

Reduced Kernel Extreme Learning Machine

Wanyu Deng, Qinghua Zheng, and Kai Zhang

Abstract. We present a fast and accurate algorithm—reduced kernel extreme learning machine (Reduced-KELM). It randomly selects a subset from given dataset, and uses $\mathcal{K}(X, \tilde{X})$ in place of $\mathcal{K}(X, X)$. The large scale kernel matrix with size of $n \times n$ is reduced to $n \times \tilde{n}$, and the time-consuming computation for inversion of kernel matrix is reduced to $O(\tilde{n}^3)$ from $O(n^3)$ where $\tilde{n} \ll n$. The experimental results show that Reduced-KELM can perform at a similar level of accuracy as KELM and at the same time being significantly faster than KELM.

1 Introduction

Kernel extreme learning machine (KELM) [1] generalize extreme learning machine [2] from explicit activation function to implicit mapping function. KELM can produce better generalization than ELM in most applications. But like other kernel machines such as support vector machine (SVM) [3], least square support vector machine(LS-SVM) [4], kernel principle component analysis(KPCA) [5]. The entire samples have to be stored in the memory. When the samples are large, the computer may run out of memory. Additionally, it is very time-consuming for kernel matrix operations such as inversion, multiplication and eigenvalue decomposition. In order to accelerate the computation efficiency, many methods have been proposed such as sequential minimal optimization (SMO) [6], Kronecker product decomposition [7], random samples [8] [9], etc. SMO is an iterative algorithm for solving the optimization problem described above. SMO breaks this problem into a series of smallest possible sub-problems, which are then solved analytically. Kronecker decomposition decomposes the large kernel matrix with size into two small matrices with

Wanyu Deng · Kai Zhang
Xi'an University of Posts & Telecommunications, China
e-mail: wanyu.deng@gmail.com, zhangkaial@163.com

Qinghua Zheng
Xi'an Jiaotong University, China
e-mail: qhzheng@mail.xjtu.edu.cn

small size, and then run eigenvalue decomposition or other corresponding operation on these two small matrix. The main idea of random sample method is selecting small subset from the entire dataset and finding one hyperspace in this subset. The method was introduced into LS-SVM and produced good performance. Motivated by this work, we will apply this method in KELM and desire to get one fast algorithm Reduced-KELM. The paper is organized as follows. Section 2 gives a brief review of the KELM. Section 3 presents the derivation of Reduced-KELM. Performance evaluation of Reduced-KELM is shown in Section 4 based on the benchmark problems from UCI repository. Conclusions and future works are highlighted in Section 5.

2 Kernel Extreme Learning Machine

Extreme learning machine (ELM) was originally proposed for the single hidden layer feedforward neural networks (SLFNs) where the hidden nodes can be any piecewise nonlinear function including additive/RBF hidden nodes, multiplicative nodes, and non-neural alike nodes [10] [11]. The output function of ELM is

$$f_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta \quad (1)$$

where $\beta = [\beta_1, \dots, \beta_L]^T$ is the vector of the output weights between the hidden layer and the output layer, and $h(x) = [h_1(x), \dots, h_L(x)]$ is the outputs of the hidden nodes with respect to the input x . $h(x)$ actually maps the data from the n -dimensional input space to the L -dimensional feature space. The minimal norm least square method was used in the original implementation of ELM[2]:

$$\beta = H^\dagger T \quad (2)$$

where H^\dagger is the Moore-Penrose generalized inverse of matrix H while H is the hidden layer output matrix:

$$H = \begin{pmatrix} h(x_1) \\ \vdots \\ h(x_n) \end{pmatrix} \quad (3)$$

one of the methods to calculate Moore-Penrose generalized inverse of a matrix is the orthogonal projection method [12]:

$$H^\dagger = H^T (HH^T)^\dagger \quad (4)$$

According to the ridge regression theory [13], one can add a positive value C to the diagonal of HH^T such that the solution is more stable and tends to produce better generalization performance:

$$f(x) = h\beta = h(x)H^T \left(\frac{I}{C} + HH^T \right)^{-1} T \quad (5)$$

Different from SVM, feature mapping $h(x)$ is usually known to users in ELM. However, if a feature mapping $h(x)$ is unknown a kernel matrix for ELM can be defined:

$$\mathcal{K}(X, X) = HH^T \quad (6)$$

Thus, the output function of kernel extreme learning machine (KELM) can be written as

$$f(x) = h(x)H^T \left(\frac{I}{C} + HH^T \right)^{-1} T = \begin{pmatrix} \mathcal{K}(x, x_1) \\ \vdots \\ \mathcal{K}(x, x_n) \end{pmatrix}^T \left(\frac{I}{C} + \mathcal{K}(X, X) \right)^{-1} T \quad (7)$$

Interestingly, KELM solution (7) is consistent to LS-SVM solution when the bias b is not used in the constraint conditions. Different from other learning algorithms [14], ELM is to minimize the training error as well as the norm of the output weights. Minimize: $\|H\beta - T\|^2$ and $\|\beta\|$

3 Reduced Kernel Extreme Learning Machine

The motivation for reduced kernel extreme learning machine (Reduced-KELM) comes from the practical objective of generating a non-linear separating surface for a large dataset which requires a small portion of the dataset for its characterization. The reasons lead to the difficulty in using nonlinear kernels on large datasets lies in twofold: 1) the kernel matrix $\mathcal{K}(X, X)$ of the entire data may be prohibitively expensive in storage and typically leads to the computer running out of memory; 2) the computing time increase dramatically with the size of samples so that it is impractical in real applications. To tackle all these difficulties, we adapt the method of random sample [2, 13] to select a small random subset $\tilde{X} = \{x_i\}_{i=1}^{\tilde{n}}$ from the original n data points $X = \{x_i\}_{i=1}^n$ with $\tilde{n} \ll n$, and use $\mathcal{K}(X, \tilde{X})$ in place of $\mathcal{K}(X, X)$ to cut problem size and computing time. This method has been applied in LS-SVM and produce good performance. Thus it is desired to be able to be applied in KELM.

Removing the regulator from the formula (7), we can get the kernel representation without regulator:

$$f(x) = \begin{pmatrix} \mathcal{K}(x, x_1) \\ \vdots \\ \mathcal{K}(x, x_N) \end{pmatrix}^T (\mathcal{K}(X, X))^{-1} T \quad (8)$$

Modifying this formulation for the reduced dataset $\tilde{X} \in R^{\tilde{n} \times m}$ with corresponding rectangular kernel matrix ,

$$f(x) = \begin{pmatrix} \mathcal{K}(x, x_1) \\ \vdots \\ \mathcal{K}(x, x_{\tilde{n}}) \end{pmatrix}^T (\mathcal{K}(\tilde{X}, X))^{-1} T \quad (9)$$

Since $\mathcal{K}(X, \tilde{X}) \in R^{n \times \tilde{n}}$ is a thin matrix, we can calculate Moore-Penrose generalized inverse by the formula $\mathcal{K}(X, \tilde{X})^{-1} = (\mathcal{K}(X, \tilde{X})^T \mathcal{K}(X, \tilde{X}))^{-1} \mathcal{K}(X, \tilde{X})^T$ to accelerate efficiency:

$$f(x) = \begin{pmatrix} \mathcal{K}(x, x_1) \\ \vdots \\ \mathcal{K}(x, x_{\tilde{n}}) \end{pmatrix}^T (\mathcal{K}(X, \tilde{X})^T \mathcal{K}(X, \tilde{X}))^{-1} \mathcal{K}(X, \tilde{X})^T T \quad (10)$$

According to the ridge regression theory [15], one can add a positive value to the diagonal of $\mathcal{K}(X, \tilde{X})^T \mathcal{K}(X, \tilde{X})$ such that the solution is more stable and tends to have better generalization performance:

$$f(x) = \begin{pmatrix} \mathcal{K}(x, x_1) \\ \vdots \\ \mathcal{K}(x, x_{\tilde{n}}) \end{pmatrix}^T \left(\frac{I}{C} + \mathcal{K}(X, \tilde{X})^T \mathcal{K}(X, \tilde{X}) \right)^{-1} \mathcal{K}(X, \tilde{X})^T T \quad (11)$$

The Reduced-KELM algorithm can be summarized as following:

Algorithm Reduced-KELM Algorithm

Given a training set $\mathfrak{X} = \{(x_k, t_k) | x_k \in R^m, t_k \in R^\ell\}_{k=1}^n$, an kernel function,

(1) Choose a random subset matrix $\tilde{X} \in R^{\tilde{n} \times m}$; Typically $\tilde{n} \ll n$.

(2) Construct rectangular kernel matrix $\mathcal{K}(X, \tilde{X}) \in R^{n \times \tilde{n}}$;

(3) Get the prediction for one new data x

$$f(x) = \begin{pmatrix} \mathcal{K}(x, x_1) \\ \vdots \\ \mathcal{K}(x, x_{\tilde{n}}) \end{pmatrix}^T \left(\frac{I}{C} + \mathcal{K}(X, \tilde{X})^T \mathcal{K}(X, \tilde{X}) \right)^{-1} \mathcal{K}(X, \tilde{X})^T T$$

4 Experimental Results

We evaluate and compare the performance of the proposed Reduced-KELM with ELM and KELM on six classification problems described in Table 1. For each problem, the results are averaged over 50 trials. The average training time, testing time, training accuracy and testing accuracy are reported.

All the simulations have been conducted in MATLAB 12 environment running on an ordinary PC with 2.6 GHZ CPU. The Gaussian RBF activation function has

Table 1 Specification of Classification Benchmark Datasets

Dataset	#Training samples	#Testing samples	#Attributes	#Classes
Segment	1500	810	18	7
Satimage	4435	2000	36	7
Shuttle	43500	14500	9	7
SkinSeg	145057	100000	4	2
Waveform	3000	2000	40	3
Madelon	1000	1600	500	2

Table 2 Parameters of Reduced-KELM, ELM, and KELM

Dataset	Reduced-KELM(\tilde{n}, C, σ)	ELM(#Nodes)	KELM(C, σ)
Segment	(400, $2^5, 2^{-1}$)	200	($2^6, 2^{-1}$)
Satimage	(700, $2^9, 2^1$)	500	($2^4, 2^{-2}$)
Shuttle	(50, $2^1 0, 2^{-1}$)	50	—
SkinSeg	(300, $2^1 0, 2^{-3}$)	300	—
Waveform	(300, $2^0, 2^0$)	400	($2^6, 2^{-1}$)
Madelon	(400, $2^2, 2^1$)	400	($2^2, 2^1$)

Table 3 The Training Time and Testing Time of Reduced-KELM, ELM, and KELM

Dataset	Reduced-KELM		ELM		KELM	
	(Training,	Testing)	(Training,	Testing)	(Training,	Testing)
Segment	0.065	0.011	0.2390	0.012	0.4673	0.043
Satimage	0.386	0.287	3.1044	0.115	17.4850	0.046
Shuttle	0.096	0.020	0.4867	0.0338	—	—
SkinSeg	2.963	0.825	27.9034	1.674	—	—
Waveform	0.082	0.023	4.676	0.0988	1.289	0.199
Madelon	0.079	0.053	0.665	0.2080	0.189	0.130

been used in the simulations of Reduced-KELM and KELM, while sigmoid function is used in ELM. The input attributes are normalized into the range [-1, 1]. Table 2 summarizes the parameter settings, while Table 3 and 4 report the results for the classification problems for each algorithm. As observed from Table 3 and 4, the training time taken by Reduced-KELM is much less than ELM and KELM while testing accuracy is better than ELM and slightly less than KELM. Take Satimage dataset as example, the training time for Reduced-KELM is 0.3867s, ELM is 3.1044s while KELM is 17.485s. Reduced-KELM is 9 times faster than ELM while 50 times faster than KELM. For the dataset Shuttle and Skin Segmentation, since the training samples are too large, the KELM will run out of memory. This shows

Table 4 The Training Accuracy and Testing Accuracy of Reduced-KELM, ELM, and KELM for Classification Problems

Dataset	Reduced-KELM		ELM		KELM	
	(Training, Testing)					
Segment	0.974	0.959	0.975	0.939	0.998	—
Satimage	0.935	0.911	0.925	0.896	1.0000	—
Shuttle	0.988	0.988	0.979	0.979	—	—
SkinSeg	0.998	0.9981	0.997	0.997	—	—
Waveform	0.867	0.85	0.895	0.838	1.00	0.85
Madelon	0.742	0.57	0.806	0.531	1.00	0.596

that Although KELM can produce the best generalization ability, it is unsuitable for large size problems.

5 Conclusion and Future Work

In this paper, a fast and accurate reduced kernel extreme learning machine (Reduced-KELM) has been developed. It uses a randomly selected subset of the data (typically much less than the original dataset) to obtain a nonlinear separating surface. Although Reduced-KELM's testing accuracy is slightly less than KELM that uses the entire data, it is tens of times faster than KELM. This is very important for massive datasets such as those in the millions scale. Reduced-KELM appears to be a very promising method for handling large problems. Additionally, Reduced-KELM is suitable to implement online sequential learning for data stream. There are some open problems for the current work such as 1) does random selection affect the prediction performance, 2) how to select the optimized subset and 3) what size the subset should be. These are our next work in the future.

References

1. Huang, G.-B., Zhou, H., Ding, X., Zhang, R.: Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 42(2), 513–529 (2012)
2. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: Extreme learning machine: Theory and applications. *Neurocomputing* 70(1-3), 489–501 (2006)
3. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, Heidelberg (1999)
4. Suykens, J.A.K., Vandewalle, J.: Least squares support vector machine classifiers. *Neural Processing Letters* 9(3), 293–300 (1999)

5. Scholkopf, B., Smola, A., Mller, K.R.: Kernel principal component analysis. In: Gerstner, W., Hasler, M., Germond, A., Nicoud, J.-D. (eds.) ICANN 1997. LNCS, vol. 1327, pp. 583–588. Springer, Heidelberg (1997)
6. Keerthi, S.S., Shevade, S.K., Bhattacharyya, C., Murthy, K.R.K.: Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Computation* 13, 637–649 (2001)
7. Wu, G., Zhang, Z., Chang, E.Y.: Kronecker factorization for speeding up kernel machines. In: SIAM Int. Conference on Data Mining (SDM), pp. 611–615 (2005)
8. Lin, K.M., Lin, C.J.: A study on reduced support vector machines. *IEEE Transactions on Neural Networks* 14(6), 1449–1459 (2003)
9. Liu, Q., He, Q., Shi, Z.-Z.: Extreme support vector machine classifier. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS (LNAI), vol. 5012, pp. 222–233. Springer, Heidelberg (2008)
10. Huang, G.B., Chen, L.: Enhanced random search based incremental extreme learning machine. *Neurocomputing* 71(16), 3460–3468 (2008)
11. Huang, G.B., Chen, L.: incremental extreme learning machine. *Neurocomputing* 70(16), 3056–3062 (2007)
12. Courrieu, P.: Fast computation of Moore-Penrose inverse matrices. *Neural Information Processing - Letters and Reviews* 8(2), 25–29 (2005)
13. Hoerl, A.E., Kennard, R.W.: Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1), 55–67 (1970)
14. Rumelhart, D.E., Hintont, G.E., Williams, R.J.: representations by back-propagating errors. *Nature*. 323(6088), 533–536 (1986)
15. Hoerl, A.E., Kennard, R.W.: Ridge regression: biased estimation for nonorthogonal problems 42(1), 80–86 (2000)

Using Positional Information in Modeling Inflorescence Discs

Malgorzata Prolejko

Abstract. The aim of this paper is the introduction to the concept of positioners. It is the new way to use the positional information while modeling compound models of plants or their parts. A specific way of compound fruit growth requires to search new solutions and further develop techniques already known. The result was obtained from decomposition of the module and applying positional information to its featured parts.

First, there is presented a mathematical description of compound objects, called geometrical method. Next, positioners are described as an extension of geometrical method. Furthermore some examples are demonstrated.

Keywords: positioner, L-system, plant growth, sunflower.

1 Introduction

Most plants have a modular structure, which allows distinction of fragments of the plant that develops in a similar way. With this observation, methods characteristic for modular objects can be used. The most popular method is L-systems [4]. There are many papers describing, expanding or modifying algorithms based on L-systems [2] [3] [5]. Most of them assume, that modules are identical and similarly develop in time. The modules in one compound can be distinguished only by their positional information.

The aim of this article is to present a new method of using positional information during modeling modular objects. The new concept called *positioners* is introduced. Its main task is to manage multiple semi-similar modules in continuous time, which is difficult in L-system methods.

Malgorzata Prolejko

Faculty of Mathematics and Computer Science, University of Warmia and Mazury, Olsztyn
e-mail: m.prolejko@matman.uwm.edu.pl

2 Geometrical Method of Modeling Modular Objects

A mathematical method of modular modeling based on L-systems is presented in [2]. The main change is that it operates in continuous time domain rather than discrete steps. Furthermore, graphic representation of modules is combined with hierarchy structure in one equation (3). Its main advantage for plant modelers is that it allows to avoid searching differential growth functions of plant parts [3]. Nevertheless, the method is limited. It uses only one kind of module, which is equivalent of L-system with one letter.

Mathematical background

Geometrical method is a mathematical notation of evolution model in time and space. It is used to describe three-dimensional shape of modules depending on time as well as their hierarchical structure. The general assumption is that the object consists of modules of the same kind, developing identically in time and space, only with delay dependent on a hierarchy level.

The set $V(t)$ consists of points belonging to a single module. It describes a space-time development in its local coordinate system. The hierarchical structure is build based on *characteristic vertices*. These are points to which child modules are assembled (Fig. 1). Characteristic vertices can move freely and their movement is described by movement matrix $\mathbf{D}(t)$ [2]. There is exactly one child module associated with each characteristic vertex. Child modules are placed according to the coordinate system of their characteristic vertices and move along with them. The number of characteristic vertices of one module is denoted by q and called *branching degree*.

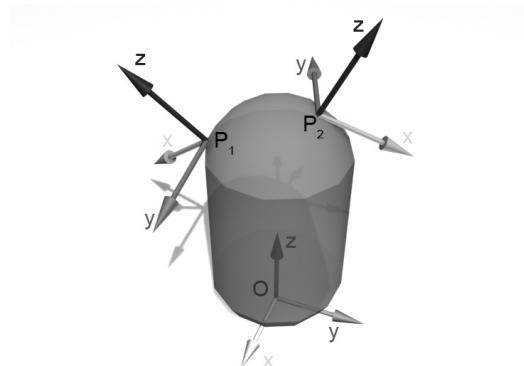


Fig. 1 Local coordinate systems of the characteristic vertices P_1 , P_2 and the module O

Let $\mathbf{p}(t)$ be the position of a vertex P in local coordinate system of the module. Position of vertex P in a parental coordinate system $\mathbf{p}'(t)$ can be calculated with the equation $\mathbf{p}'(t) = \mathbf{D}(t)\mathbf{p}(t)$. Generalizing it on all vertices, a description of the child module is obtained [2]:

$$\mathbf{V}'(t) = \mathbf{D}(t)\mathbf{V}(t). \quad (1)$$

Not all modules develop simultaneously. For a given plant species a constant delay T (*modulus*) can be defined. It determines the time delay after which child modules start to develop. Then the child module can be described as follows (2):

$$\mathbf{V}'(t) = \mathbf{D}(t)\mathbf{V}(t - T) \quad (2)$$

Putting all modules in hierarchy together description of the whole object $Z(t)$ is obtained:

$$Z(t) = V(t) \cup \bigcup_{i=1}^k \bigcup_{\hat{g}^i \in \mathbf{G}_i} \mathbf{D}_{\hat{g}^i}(t)V(t - iT). \quad (3)$$

The equation (3) is the union of root module $V(t)$ and unions of the all child modules in hierarchy structure. i stands for level of the hierarchy and \mathbf{G}_i is the set of all possible combination of ancestors indexes in level i .

In the basic approach (3), there is no time limit, which means that the plant develops infinitely. The maximum level number k can be given, so that none module with that hierarchy level produce children. The development of a single module is finished in time T_d .

$$t_s = k * T + T_d \quad (4)$$

Limiting the number of levels to a preset value determines static state after t_s , when all modules do not develop (4) [2].

Possible extensions

The primary assumption of the identity of the modules not always suits for modeling process. We can distinguish a group of cases in which the development of the older modules differs from the others. Geometric method allows to use positional information [5] to formulate a set $V_i(t)$. Modification of the coordinate systems can also take place.

Assumptions and definitions necessary to be declared before start to produce compound objects are shown below.

- $V(t)$ – single module space-time development,
- q – branching degree,
- $\mathbf{D}_p(t)$ – set of movement matrices for every characteristic vertex,
- T – time delay,
- k – hierarchy levels number,
- T_d – maturation time of the module $V(T)$.

3 The Concept of Positioners

The concept of *positioners* is an extension of the geometrical method, which was introduced as a mean of constructing flower models with spiral phyllotaxis. It allows to differentiate parts of module (e.g. single flower) based on its hierarchy index number [2].

Modeling a sunflower inflorescence disc with basic graphical method provides the unrealistic results. The reason for that is the identical module development being one of the bases of the aforementioned method. It strictly associates the position of flower module on a disc ray with its development stage. As a result, all flower-modules are mature only on a disc edge.

One can exclude the movement on the disc from the development process. Then flower-module only ripens and inflorescence arrangement must be done separately. Moreover, information about hierarchy is contained in the arrangement structure, so its impossible to delay growth of the flowers in the center of the disc. Positioners allow to differentiate rather than exclude the movement from the development process in order to keep positional information for entire module.

There are two examples to use positioners described in this paper. The first one shows how positioner determines position of flower on inflorescence disc. The second one shows how multiple plant species can be modeled with one basic module.

3.1 The Positioner Characteristic

Separation of movement from module development motivates creation of separator fundamental to decomposition. In basic geometrical method, positional information was applied with function F_i [2] (5).

$$V_i(t) = F_i(V(t)) \quad (5)$$

This function is applied to all points from the set $V(t)$. When the decomposition (6) is conducted, new local coordinate systems must be selected. The ones that will be modified, are separated into category of *positioner*.

$$V(t) = \bigcup_{j=1}^n v_j(t) \quad (6)$$

Decomposition of the set $V(t)$ into n sub-modules $v_j(t)$ ($j = 1, \dots, n$) is shown in Fig. 2. A single module consists of a set of points, dedicated local coordinate system and characteristic vertices with their own coordinate systems (Fig. 2a). After decomposition module is divided into new sets of points (Fig. 2b). The coordinate system of the *Sphere SXYZ* is positioner. Consequently, element *Sphere* has positioner status.

Inner hierarchy of a single module introduced in Fig. 2b. Rectangle indicates coordinate system of a single element. The element which is highlighted (*Sphere*) is a positioner with its coordinate system singled out.

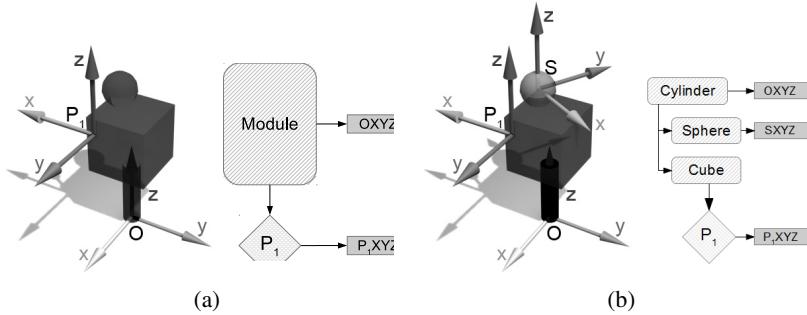


Fig. 2 Objects in module: a) basic method, b) with positioner

Part of the module which gains positioner status is denoted by $\tilde{v}_j(t)$ and its index $j \in \{1, \dots, n\}$ is added to the positioner index set P . Fragments with indexes in a complementary set $P' = \{1, \dots, n\} - P$ are not modified by the function F_i . The whole module is described by equation (7).

$$V_i(t) = \bigcup_{j \in P'} v_j(t) \cup \bigcup_{j \in P} F_i(\tilde{v}_j(t)) \quad (7)$$

Positioner movement in coordinate system of the module is described by a displacement matrix $\mathbf{H}(t)$ (8).

$$\mathbf{H}(t) = \begin{bmatrix} r_{1x}(t) & r_{2x}(t) & r_{3x}(t) & d_x(t) \\ r_{1y}(t) & r_{2y}(t) & r_{3y}(t) & d_y(t) \\ r_{1z}(t) & r_{2z}(t) & r_{3z}(t) & d_z(t) \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Each positioner has its own matrix $\mathbf{H}(t)$. Therefore for the purpose of the whole module a set of matrices is designed $\{\mathbf{H}_j(t)\}$, where $j \in P$. These matrices must be taken into account while defining sets $\tilde{v}_j(t)$ (9).

$$\tilde{v}_j(t) = \mathbf{H}_j(t) \tilde{v}'_j(t) \quad (9)$$

Function F_i refers to whichever transforming function. For the purposes of this paper, a composition of two scaling functions are taken into account. The first one (Ψ) scales the size and second (Φ) scales position of positioner and its children. Both are index dependent and time independent. Scaling matrices $\mathbf{S}(i)$ (10) [1] are used to define functions Ψ and Φ .

$$\mathbf{S}(i) = \begin{bmatrix} s_x(i) & 0 & 0 & 0 \\ 0 & s_y(i) & 0 & 0 \\ 0 & 0 & s_z(i) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (10)$$

Function Φ , which scales the size of the object is typical (11).

$$\Psi(i) = \mathbf{S}_\Psi(i) \quad (11)$$

Function $\Phi(i)$ is defined by the scaling matrix $\mathbf{S}_\Phi(i)$, which is applied to the displacement matrix $\mathbf{H}_j(t)$ (12).

$$\Phi(i) = \mathbf{S}_\Phi(i) \mathbf{H}_j(t) \mathbf{S}_\Phi^{-1}(i) \quad (12)$$

Function F_i assemblies scalings (12) and (11).

$$F_i(\tilde{v}_j(t)) = \mathbf{S}_\Phi(i) \mathbf{H}(t) \mathbf{S}_\Phi^{-1}(i) \mathbf{S}_\Psi(i) \tilde{v}'_j(t) \quad (13)$$

After decomposition of the module (7) with inclusion of the modifying function F_i , the entire object $Z(t)$ (3) can be described by equation (14).

$$Z(t) = V(t) \cup \bigcup_{i=1}^k \bigcup_{\hat{g}^i \in \mathbf{G}_i} \mathbf{D}_{\hat{g}^i}(t) \left(\bigcup_{j \in P'} v_j(t - iT) \cup \bigcup_{j \in P} F_i(\tilde{v}_j(t - iT)) \right) \quad (14)$$

Overall, the space-time development of the entire object can be easily described with the notation obtained from geometrical method. Every element of the equation (14) has simple representation in any 3d design program, thus implementation of the objects is intuitive.

3.2 The Use of Positioners to Model a Inflorescence Disc

Various stages of the single disc flower development are position independent. Defining positioners helps to separate a single ray flower development from inflorescence modeling (Fig. 3) during modeling process. The use of the positioners to create inflorescence disc are described in this subsection.

Various configurations of positioner scaling based on one module structure is shown in Fig. 4. A module is composed with animated sphere *Sphere*, cylinder *Cylinder* and three Dummy objects: *Parent*, *Child* and *Positioner*. Maturation time of the module is $T_d = 500$ frames. Animated texture allows to determine the state of development in current frame - from light gray in early stages to dark gray when module is mature. The *Cylinder* has variable length and is positioned along *OY* axis in the coordinate system of the module. The *Positioner* is strictly connected with the center of the sphere, in order to be observed the movement of the whole module.

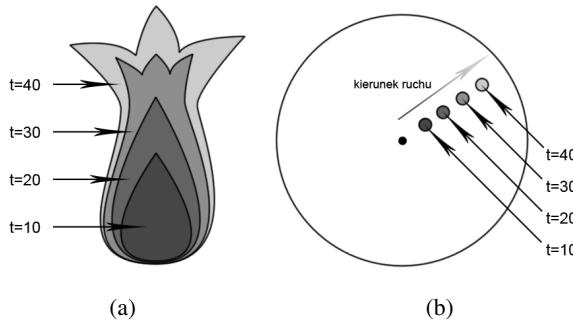


Fig. 3 Separation animation of the module from its movement on the disc: a) module development; b) movement of the module along disc ray

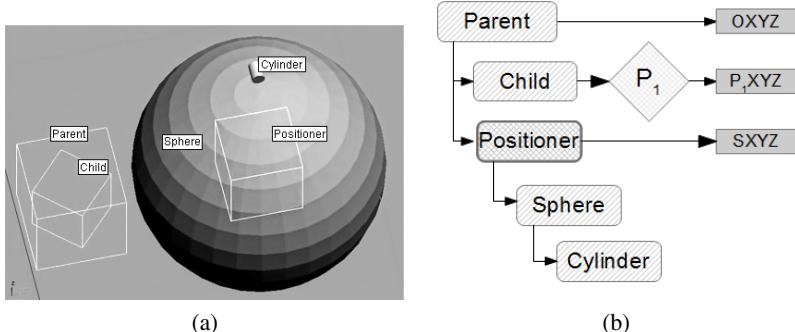


Fig. 4 Module prepared to simulation: a) elements, b) hierarchy

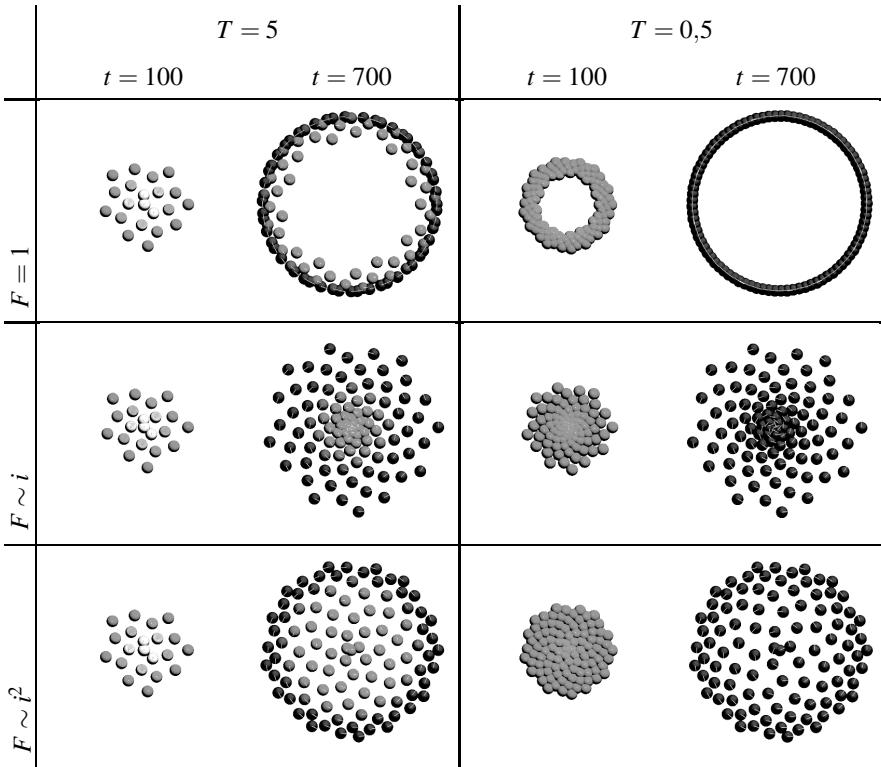
There were six simulations performed on the module (Fig. 4). All of them had the same hierarchy levels number $k = 100$. Simulation parameters with visualizations of 100th and 700th frame are presented in Table 1. All three scaling functions Φ depend only on the index of the module. There were no Ψ functions included.

Two different modulus T apportion columns of the Table 1. In rows, different scaling functions Φ are applied. In every table cell, there are two figures. The left ones are obtained from frame 100th and the right ones from frames 700th.

$$n = \min(\lfloor frame/T \rfloor + 1, k) \quad (15)$$

Different modulus T affects the time the object matures and the number of produced modules n (15). Discs in the left column have 21 modules in 100th frame and in all other figures there are all 100 modules. The state is constant when $t > t_s$ (4) and all the modules in right column are mature (frame $t = 700$). Color of a single module determines its development stage. The bigger the modulus, the more contrast of the modules color are in the object.

Table 1 Table of disc views with different simulation parameters. Left images are acquired for frame 100 and the right ones are for 700.



The scaling function has an effect in arrangement of modules in the disc. While the divergence angle is constant and equal $34/89$ ($137^\circ 31'$) [6], scaling applies only to the distance from center of the disc. In the case when scaling does not take place (row 1 of Table 1), similarly to the basic geometrical method, development stage of a single module depends on its position on the disc ray. When all modules are mature, they all are placed on the disc edge. After linear scaling, dense group of modules appears in the center of the disc (row 2 of Table 1). Quadratic scaling causes the arrangement of modules to be quite uniform. This is a more satisfying and awaited effect.

3.3 The Use of Positioners to Differentiate Parts of the Module

Any object in the module hierarchy can be picked as a positioner during modeling process. This positioner, with its children, will be differentiated with scale factor from other elements in the module. This can be observed on the example of the ear.

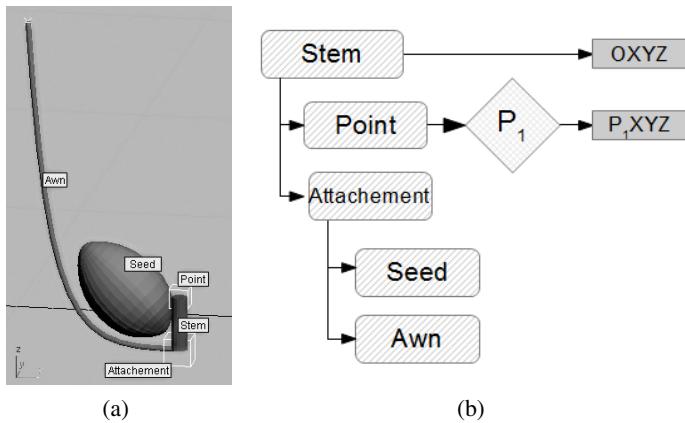


Fig. 5 Seed of the ear prepared for simulation: a) elements of the module, b) hierarchical structure

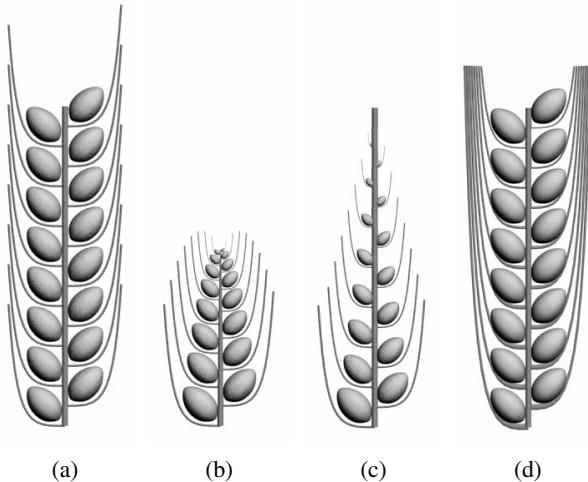


Fig. 6 Examples of mature ears (frame 150) created with geometrical method: a) without positional information, b) with scaling according to positional information, c) and d) with use of positioners

Fig. 5 presents the module structure, which was used as a study example. The module consists of the stem portion *Stem* which is the root of the hierarchy tree. The characteristic vertex is defined by object *Point*. Elements *Seed* and *Awn* are attached to *Attachement*. This kind of structure allows to pick positioners in several ways. Fig. 6 shows final state of four kinds of ears. Each of them consists of 16 modules.

First ear (Fig. 6a) was the result of the basic geometrical method where every module has the same size. It resembles rye. Fig. 6b shows effect of linear scaling module according to its positional information. The higher the level, the smaller the module is.

Fig. 6c and 6d show ears with positioners scaled. *Attachement* is a positioner in the ear on figure 6c and positional information determines the scale of *Seed* and *Awn*. In the last example 6d, there is *Awn* selected as positioner. It was possible to fit scale factor so that all awns end equally as in ear of barley. All seeds of that ear have the same size like there was no scaling 6a.

4 Conclusion

To sum up, the use of L-systems is problematic in view of separation between graphical representation and structure notation [4]. Additionally, it is difficult to find proper differential growth function when one wants to use continuous time space [3]. In the contrary, geometrical method allows to operate freely in continuous time space and it is simple to implement in graphical programs. Positional information can be applied with minimum effort, achieving satisfying effects.

References

1. Foley, J., van Dam, A., Feiner, S.K., Hughes, J.F., Philips, R.L.: *Introduction to Computer Graphics*. Addison-Wesley, MA (1990)
2. Izdebska-Prolejko, M., Stêpieñ C.: The modeling of modular plants using positioners on the example of sunflower. MSc Thesis, Warsaw Technical University, Warsaw, in polish (2010)
3. Prusinkiewicz, P., Hammel, M., Mjolsness, E.: Animation of Plant Developement. In: *Proceedings of SIGGRAPH*, vol. 93, pp. 351–360 (1993)
4. Prusinkiewicz, P., Lindenmayer, A.: *The Algorhythmic Beauty of Plants*. Springer, New York (1990)
5. Prusinkiewicz, P., Muendermann, L., Karwowski, R., Lane, B.: The Use of Positional Information in the Modeling of Plants. In: *Proceedings of SIGGRAPH*, pp. 289–300 (2001)
6. Strasburger, E.: *Botany*. Państwowe Wydawnictwo Rolnicze i Leśne, Warsaw, pp. 165–168 (1972) (in polish)

A New Method for Random Initialization of the EM Algorithm for Multivariate Gaussian Mixture Learning

Wojciech Kwedlo

Abstract. In the paper a new method for random initialization of the EM algorithm for multivariate Gaussian mixture models is proposed. In the method booth mean vector and covariance matrix of a mixture component are initialized randomly. The mean vector of the component is initialized by the feature vector, selected from a randomly chosen set of candidate feature vectors, located farthest from already initialized mixture components as measured by the Mahalanobis distance. In the experiments the EM algorithm was applied to the clustering problem. Our approach was compared to three well known EM initialization methods. The results of the experiments, performed on synthetic datasets, generated from the Gaussian mixtures with the varying degree of overlap between clusters, indicate that our method outperforms three others.

1 Introduction

Mixture models [10] are very useful tool, widely applied in pattern recognition to model complex probability distributions. A finite mixture model $p(\mathbf{x}|\Theta)$ can be expressed by a weighted sum of $K > 1$ components:

$$p(\mathbf{x}|\Theta) = \sum_{m=1}^K \alpha_m p_m(\mathbf{x}|\theta_m), \quad (1)$$

where α_m is m -th mixing proportion and p_m is the probability density function of the k th component. In (1) θ_m is the set of parameters defining the m th component and $\Theta = \{\theta_1, \theta_2, \dots, \theta_K, \alpha_1, \alpha_2, \dots, \alpha_K\}$ is the complete set of the parameters needed

Wojciech Kwedlo

Faculty of Computer Science, Bialystok University of Technology

Wiejska 45a, 15-351 Białystok, Poland

e-mail: w.kwedlo@pb.edu.pl

to define the mixture. The functional form of p_m is assumed to be known; Θ is unknown and has to be estimated. The mixing proportions satisfy the following conditions: $\alpha_m > 0$, $m = 1, \dots, K$ and $\sum_{m=1}^K \alpha_m = 1$. The number of components K is either known a priori or has to be determined during the mixture learning process. In the paper we assume, that K is known.

In the paper we consider the most widely used class of mixture models called Gaussian mixture models (GMM), in which the probability density function of the m th component is given by:

$$p_m(\mathbf{x}|\theta_m) = \frac{1}{(2\pi)^{d/2} |\Sigma_m|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mu_m)^T \Sigma_m^{-1} (\mathbf{x} - \mu_m)\right), \quad (2)$$

where μ_m and Σ_m denote the mean vector and covariance matrix, respectively, $|\cdot|$ denotes a determinant of a matrix and d is the dimension of the feature space. The set parameters of the m th component is $\theta_m = \{\mu_m, \Sigma_m\}$. Thus, for the GMM Θ is defined by: $\Theta = \{\mu_1, \Sigma_1, \dots, \mu_K, \Sigma_K, \alpha_1, \dots, \alpha_K\}$.

Estimation of the parameters of the GMM can be performed using the maximum likelihood approach. Given a set of independent and identically distributed feature vectors $X = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N\}$, called the training set, the loglikelihood corresponding to the K -component GMM is given by:

$$\log p(X|\Theta) = \log \prod_{i=1}^N p(\mathbf{x}^i|\Theta) = \sum_{i=1}^N \log \sum_{m=1}^K \alpha_m p_m(\mathbf{x}^i|\theta_m). \quad (3)$$

The maximum likelihood estimate of the parameters is given by: $\Theta_{ML} = \text{argmax}_{\Theta} \{\log p(X|\Theta)\}$. It is well known that the solution of this maximization problem cannot be obtained in closed form (e.g. [1]). Thus, a numerical optimization method has to be employed to find it.

Because of simple implementation, the EM algorithm [3] is the most popular tool for maximum likelihood estimation of the parameters of the GMM. However, this procedure is highly sensitive to initialization and easily gets trapped in local optimum of (1). Therefore, the quality of the final solution is strongly dependent on the initial guess of the model parameters. This problem can be somehow alleviated by performing multiple runs of the algorithm, each run started from different random initial conditions, and returning the result with the highest $\log p(X|\Theta)$. In an another approach to the initialization issue the EM algorithm is combined with some global optimization method e.g. genetic algorithm [12] or a particle swarm optimizer [2]. However global optimizers have high computational demands and this approach is limited to moderately sized datasets.

Clustering [4, 15] is the most important application of GMMs. In this application each feature vector is assumed to be generated from one K mixture components. The goal of clustering is to identify, for each feature vector, the mixture component from which it was generated. If we know the mixture parameters Θ , we can achieve this by allocating a feature vector \mathbf{x}^i to a cluster (mixture component) with the

highest posterior probability. Using Bayes formula this probability for the mixture component m can be expressed as:

$$h_m(\mathbf{x}^i) = \frac{\alpha_m p_m(\mathbf{x}^i | \theta_m)}{p(\mathbf{x}^i | \Theta)}. \quad (4)$$

Maximization of (4) is equivalent to finding the mixture index m with the highest value $\alpha_m p_m(\mathbf{x}^i | \theta_m)$.

The standard method for random initialization of the EM algorithm for GMMs [10] initializes a mean vector μ_m with a randomly chosen feature vector. After initialization of the mean vectors, the feature vectors are clustered by assigning a feature vector to the group represented by the closest mean vector. Next, the covariance matrix and the mixing proportion of each group are used as initial estimates of parameters of a mixture component. A variant of this method uses the K -means algorithm [11] to initialize component means.

In the paper we propose a new method for initialization of the EM algorithm for GMMs. Our method, while retaining the probabilistic nature of random initialization, tries to initialize component means by feature vectors located far away from already initialized components.

The remainder of the paper is organized as follows. In the next section the EM algorithm for GMMs is described. Section 4 presents our approach to initialization. Section 5, shows the results of computational experiments in which our approach was compared to three other initialization methods. The last section concludes the paper.

2 EM Algorithm for Gaussian Mixture Models

The most popular approach for maximizing (3) is the EM algorithm. It is an iterative algorithm, which, starting from initial guess of a parameters $\Theta^{(0)}$, generates a sequence of estimations $\Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(j)}, \dots$, with increasing loglikelihood (i.e., $\log p(X|\Theta^{(j)}) > \log p(X|\Theta^{(j-1)})$). Each iteration j of the algorithm consists of two steps called expectation step (E-step) and maximization step (M-step). For the GMM these steps are defined as follows [13]:

- E-Step: Given the set of mixture parameters $\Theta^{(j-1)}$ from the previous iteration, the posterior probability that a sample \mathbf{x}^i was generated from m -th component is computed as:

$$h_m^{(j)}(\mathbf{x}^i) = \frac{\alpha_m p_m(\mathbf{x}^i | \theta_m^{(j-1)})}{\sum_{k=1}^K \alpha_k p_k(\mathbf{x}^i | \theta_k^{(j-1)})}, \quad (5)$$

where $\theta_m^{(j-1)}$ and $\theta_k^{(j-1)}$ denote parameters of components m and k , in the iteration $j - 1$, respectively.

- M-Step: Given the posterior probabilities $h_m^{(j)}(\mathbf{x}^i)$ the set of parameters $\Theta^{(j)}$ is calculated as:

$$\alpha_m^{(j)} = \frac{1}{N} \sum_{i=1}^N h_m^{(j)}(\mathbf{x}^i) \quad (6)$$

$$\mu_m^{(j)} = \frac{\sum_{i=1}^N h_m^{(j)}(\mathbf{x}^i) * \mathbf{x}^i}{\sum_{i=1}^N h_m^{(j)}(\mathbf{x}^i)} \quad (7)$$

$$\Sigma_m^{(j)} = \frac{\sum_{i=1}^N h_m^{(j)}(\mathbf{x}^i)(\mathbf{x}^i - \mu_m^{(j)})(\mathbf{x}^i - \mu_m^{(j)})^T}{\sum_{i=1}^N h_m^{(j)}(\mathbf{x}^i)} \quad (8)$$

The E-Steps and M-Steps alternate until some convergence criterion is met.

3 Description of Our Method

3.1 Random Generation of Component Means

The aim of our approach to initialization of the EM algorithm for GMM was to correct the shortcoming of pure random initialization illustrated on Figure 1. This Figure shows the stage of initialization algorithm for three-component mixture at which two component (labeled as ‘1st component’ and ‘2nd component’) means were correctly initialized by one of feature vectors marked \square . To increase chances for discovering the optimal solution it would be beneficial to initialize the mean of the third component by one of the feature vectors labeled \triangle . However, in pure random initialization method all the feature vectors, including the vectors labeled \square , have the same probability of being selected as the means of the third component. In that situation the selection of one of vectors marked by \square may lead to suboptimal solution.

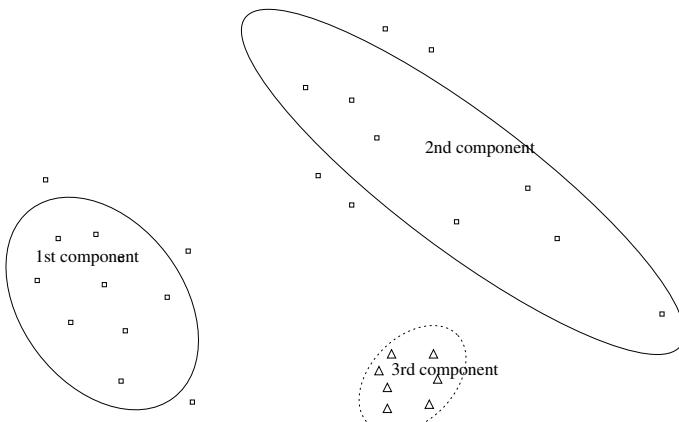


Fig. 1 An example three-component mixture illustrating the motivation for our method

To correct this shortcoming our method initializes component means and covariances incrementally. When selecting new feature vectors as component means it tries to disregard the vectors located closely to already initialized components. Our method has a probabilistic nature, which allows it to benefit from multiple runs of the EM algorithm. The initialization procedure can be described by the following steps:

1. Choose the mean of the first component $\mu_1^{(0)}$ as a random feature vector. Generate randomly a covariance matrix $\Sigma_1^{(0)}$. Set $m = 2$.
2. Choose t random unique feature vectors from the remaining (not yet selected as component means) elements of X . Denote by $X_r = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$ the set of the chosen vectors. For each $\mathbf{x}_i \in X_r$ compute the minimal squared Mahalanobis distance to the already chosen component means $\mu_1^{(0)}, \dots, \mu_{m-1}^{(0)}$:

$$dmin^2(\mathbf{x}_i) = \min_{j=1, \dots, m-1} d_M^2(\mu_j^{(0)}, \Sigma_j^{(0)}, \mathbf{x}_i), \quad (9)$$

where $d_M^2(\mu_j^{(0)}, \Sigma_j^{(0)}, \mathbf{x}_i)$ is given by:

$$d_M^2(\mu_j^{(0)}, \Sigma_j^{(0)}, \mathbf{x}_i) = (\mu_j^{(0)} - \mathbf{x}_i) \left(\Sigma_j^{(0)} \right)^{-1} (\mu_j^{(0)} - \mathbf{x}_i)^T \quad (10)$$

3. Select as the $m - th$ component mean:

$$\mu_m^{(0)} = \underset{\mathbf{x}_i \in X_r}{\operatorname{argmax}} dmin^2(\mathbf{x}_i) \quad (11)$$

4. Generate randomly a covariance matrix $\Sigma_m^{(0)}$
5. $m = m + 1$ if $m < K$ then goto Step 2. Otherwise terminate the algorithm.

3.2 Random Generation of Component Covariances

To generate randomly a covariance matrix $\Sigma_m^{(0)}$, we used the eigenvalue decomposition. Since a covariance matrix of a non-degenerate multivariate normal distribution is positive definite, it can be expressed as:

$$\Sigma_m^{(0)} = Q_m \Lambda_m Q_m^T, \quad (12)$$

where Λ_m is a diagonal matrix of eigenvalues with all diagonal entries positive and Q_m is an orthogonal matrix of eigenvectors. In our method we first generate eigenvalues. The eigenvalues are generated randomly with the following two restrictions: a) the relation of the largest eigenvalue to the smallest eigenvalue cannot be greater than 10, b) sum of all eigenvalues should be equal $\frac{1}{10dK} \operatorname{tr}(\Sigma)$, where tr is the trace

of a matrix and Σ is the covariance matrix of the training set X . Next the orthogonal matrix Q_m is obtained by generating randomly a square $d \times d$ matrix and performing its QR decomposition [5]. The covariance matrix $\Sigma_m^{(0)}$ is then obtained using (12).

4 Experimental Results

In this section the results of the initial computational experiments, in which the EM algorithm was applied to the clustering problem, are presented. We compared our method to two other initialization methods: random initialization and initialization by the K -means algorithm. In addition, a greedy initialization method proposed in [14] was used in comparison. When using the initialization by the K -means algorithm, the centroids were initialized by randomly chosen feature vectors.

In the experiments we used a generator recently proposed by [9], which generates randomly Gaussian mixtures according to the user-defined overlap characteristic. The overlap ω_{ij} between two clusters i and j is defined as the sum of two misclassification probabilities $\omega_{j|i}$ and $\omega_{i|j}$, where:

$\omega_{j|i} = \Pr[\alpha_i p(\mathbf{x}|\mu_i, \Sigma_i) < \alpha_j p(\mathbf{x}|\mu_j, \Sigma_j) | \mathbf{x} \sim \mathcal{N}(\mu_i, \Sigma_i)]$, and similarly $\omega_{i|j} = \Pr[\alpha_j p(\mathbf{x}|\mu_j, \Sigma_j) < \alpha_i p(\mathbf{x}|\mu_i, \Sigma_i) | \mathbf{x} \sim \mathcal{N}(\mu_j, \Sigma_j)]$, where \Pr denotes the probability of an event.

The input of the generator [9] was the parameter $\bar{\omega}$ expressing the average pairwise overlap between clusters. In our experiments the number of components K was fixed at 20. We generated mixtures with dimension $d \in \{2, 5, 10\}$. For each dimension we used $\bar{\omega} \in \{0.0001, 0.0002, 0.0005, 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05, 0.1\}$. Figure 2 shows example training sets simulated from mixtures obtained from the generator for different values of $\bar{\omega}$.

We used the adjusted Rand index (ARI) [7] to compare partitions of data discovered by the clustering algorithms with the original partitions (we knew them because we used synthetic datasets drawn by a random generator, which allowed us to track the source (i.e., mixture component) of each feature vector). The ARI takes the value between -1 and 1. The expected value of ARI in case of randomly generated partitions is 0. A higher value of ARI indicates a higher similarity between partitions; a maximum value of 1 means, that two partitions are identical.

The feature vectors were clustered according to the maximum posterior probability, as described in Section 1. Since in this experiment the true mixture parameters were available, we also performed clustering using them.

The experimental protocol was as follows. For every combination of d and $\bar{\omega}$ 50 different random mixtures was generated. For each mixture a single dataset consisting of 6000 feature vectors was realized. To assure a fair comparison, each of three random initialization methods was allocated equal CPU time. The fourth, greedy method is a deterministic one and was run only once.

In the allocated time, each of three random methods was used to initialize multiple EM runs, and the result of the best run (with the highest $\log p(X|\Theta)$) was

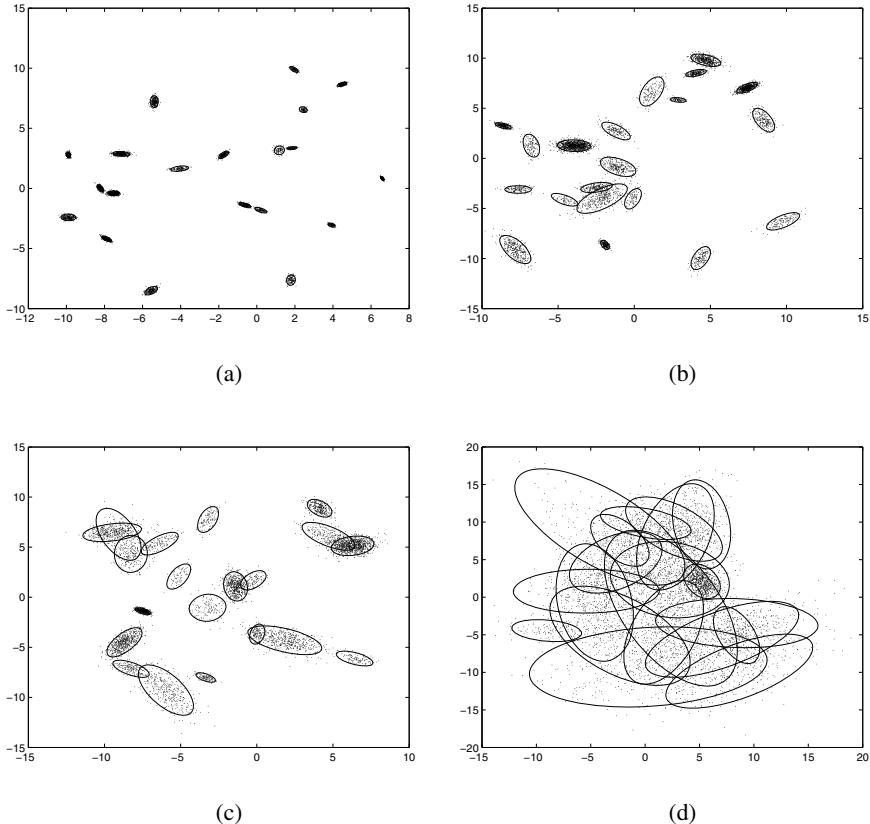


Fig. 2 Two-dimensional training sets simulated from 20-component mixtures with (a) $\bar{\omega} = 0.0001$, (b) $\bar{\omega} = 0.001$, (c) $\bar{\omega} = 0.01$, (d) $\bar{\omega} = 0.1$

considered as the result of the method. The EM algorithm was terminated when the relative improvement in loglikelihood between two consecutive iterations was smaller than $1e-5$.

Figure 3a shows the obtained values of ARI, (averaged over 50 different mixtures) when clustering was performed on the basis of 'true' parameters. As expected, whereas for clusters with very small overlap ARI close to 1 could be obtained, an increase of overlap between clusters led to lower values of ARI.

The average ARI values obtained for the true mixture parameters were used as the baseline for comparison of four estimation methods. The results concerning these methods are shown on Figures 3b, 3c, 3d. The result of each method is shown as a % error relative true mixture parameters. The % error of the method A is computed as

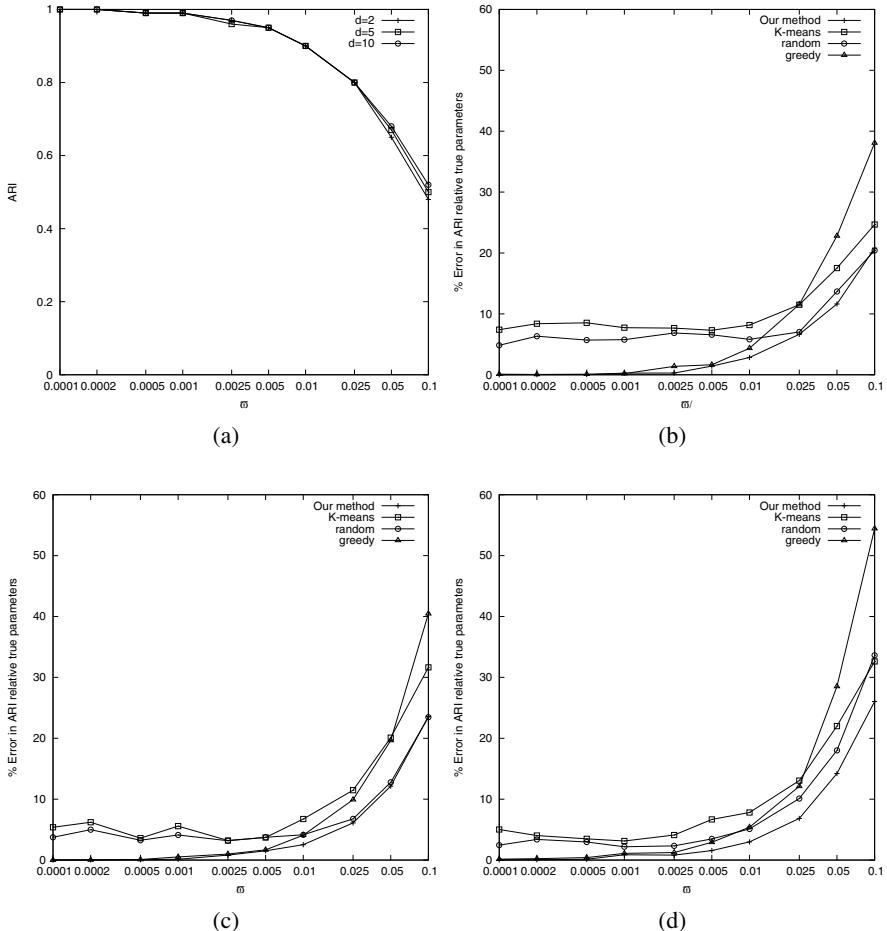


Fig. 3 (a) The average ARI values obtained for clustering using true mixture parameters. % Error in ARI relative true mixture parameters for (b) $d = 2$, (c) $d = 5$, (d) $d = 10$.

$(\text{ARI}_T - \text{ARI}_A)/\text{ARI}_T * 100$, where ARI_T is the average (over 50 different mixtures) ARI obtained using true mixture parameters and ARI_A is average ARI obtained using mixture parameters estimated by the method A .

The results from Figures 3b – 3d are summarized in Table 1, which shows the results averaged over 10 different values of $\bar{\omega}$ separately for each dimension d . The last row of the table shows the total average result for each of compared methods.

The results indicate, that our method outperforms random and K -means initialization, especially for well separated clusters. The results of greedy method are on par with our method for well separated clusters; however the greedy method does not work well, when the overlap between clusters is high.

Table 1 The average error in ARI relative known mixture parameters

d	our method	K -means	random	greedy
2	4.39	10.89	8.30	8.02
5	4.69	9.77	7.02	7.74
10	5.36	10.19	8.37	10.65
Total	4.81	10.28	7.89	8.80

5 Conclusions and Future Work

The initial experimental results allow us to conclude, that our approach achieves better results than the other three EM initialization methods. Several directions of future work exist. Although the comparison on synthetic data allowed us to compare the ARI achieved by the different variants of the EM algorithm with theoretical maximum obtained using true (not estimated) mixture parameters, we are going to test our method on real-life datasets. We also plan to apply our method to classification problems and use it to model class-conditional densities in discriminant analysis. This approach is known as Mixture Discriminant Analysis [6]. We also plan to combine our version of the EM algorithm, with a global optimization method, e.g. differential evolution, similarly as we had done with the K -means algorithm in [8].

Acknowledgements. This work was supported by the grant S/WI/2/13 from Bialystok University of Technology.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
2. Caglar, A., Aksoy, S., Arikan, O.: Maximum likelihood estimation of Gaussian mixture models using stochastic search. *Pattern Recognit.* 45(7), 2804–2816 (2012)
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B* 39(1), 1–38 (1977)
4. Everitt, B.S., Landau, S., Leese, M.: Cluster Analysis. Arnold, London (2001)
5. Golub, G.H., van Loan, C.F.: Matrix Computations. Johns Hopkins, Baltimore (1996)
6. Hastie, T., Tibshirani, R.: Discriminant analysis by Gaussian mixtures. *J. Royal Stat. Soc. Ser. B* 58(1), 155–176 (1996)
7. Hubert, L., Arabie, P.: Comparing partitions. *J. Classif.* 2(1), 193–218 (1985)
8. Kwedlo, W.: A clustering method combining differential evolution with the K-means algorithm. *Pattern Recognit. Lett.* 32(12), 1613–1621 (2011)
9. Maitra, R., Melnykov, V.: Simulating data to study performance of finite mixture modeling and clustering algorithms. *J. Comput. Graph. Stat.* 19(2), 354–376 (2010)
10. McLachlan, G., Peel, D.: Finite Mixture Models. Wiley, New York (2000)

11. McQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297 (1967)
12. Pernkopf, F., Bouchaffra, D.: Genetic-based EM algorithm for learning Gaussian mixture models. *IEEE Trans. Pattern Analysis Mach. Intell.* 27(8), 1344–1348 (2005)
13. Redner, R.A., Walker, H.F.: Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* 26(2), 195–239 (1984)
14. Verbeek, J.J., Vlassis, N., Kröse, B.: Efficient greedy learning of Gaussian mixture models. *Neural Comput.* 15(2), 469–485 (2003)
15. Xu, R., Wunsch, D.: Survey of clustering algorithms. *IEEE Trans. Neural Netw.* 16(3), 645–678 (2005)

Heuristic Optimization Algorithms for a Tree-Based Image Dissimilarity Measure

Bartłomiej Zieliński and Marcin Iwanowski

Abstract. In this paper, we present an application of three heuristic optimization algorithms to computing tree-based image dissimilarity. Genetic algorithm, particle swarm optimization and simulated annealing have been applied to optimize a black-box function which aims to determine a difference between two trees, constructed upon binary images. Presented results show that the particle swarm optimization achieved the best results. Both PSO and the simulated annealing outperformed the genetic algorithm. We also draw conclusions on parameter adjustment for the considered methods.

1 Introduction

Heuristic optimization methods have constituted an important share in optimization studies recently. The interest in this area has been intensified by successful research, which has yielded many efficient algorithms. These, often biologically inspired, solutions cope with problems, which standard approaches fail to—or simply cannot—handle. Namely, heuristic methods effectively deal with nonlinear, multiple optima functions, with no derivative information, for both continuous and discrete domains. Usually, black-box functions are challenging, as they come within mentioned classes. There is a constant drive for deeper comprehension and further improvement of this type of methods. Both external, computational needs and bottom-up algorithm examination appoint directions of advancement in the field of optimization. In this paper, we evaluate three heuristic optimization algorithms on a black-box function, which calculates dissimilarity of images. Measuring the dissimilarity is based on tree representation of binary images, where difference between

Bartłomiej Zieliński · Marcin Iwanowski

Institute of Control and Industrial Electronics,
Warsaw University of Technology, Warsaw, Poland
ul. Koszykowa 75, 00-662 Warszawa, Poland

e-mail: {bartłomiej.zielinski, marcin.iwanowski}@ee.pw.edu.pl

images is determined by comparing the trees constructed upon the images. In the current study, genetic algorithm, particle swarm optimization and simulated annealing are taken into consideration.

The paper is organized as follows. The next section describes a tree-based image dissimilarity measure which underlies the experiments. In Section 3 we briefly present examined optimization algorithms. Experimental results are shown and discussed in Section 4. Conclusions are drawn in Section 5.

2 Tree-Based Image Dissimilarity

Heuristic optimization is applied in the current study to find best match of two trees created upon binary images. This matching is necessary to compute the dissimilarity between the trees in view of evaluating the dissimilarity of two underlying images.

We take advantage of dissimilarity measures introduced in [15], [16]. In this paper, we apply tree matching optimization for trees constructed upon binary images. The parameters of the tree depend on pixel position in an image.

A binary image is usually perceived as a rectangular matrix of two-valued scalars. Let $S = \{p: I(p) = u\}$, where p stands for pixel coordinates of an image I and u is a value of foreground pixels. In order to compute the tree for a given binary image the following algorithm is applied. In successive steps, subsets of the input set of foreground pixels are created. Until the desired tree level is not reached, sets are recursively determined. For all the sets, characteristic elements are calculated. The tree is constructed by connecting the characteristic elements. As they are bound with sets of pixels, the elements remain in parent-child relationship as well and a tree arises by linking the elements which fulfil the relation. In order to apply the algorithm, one has to provide the input set S , a function f which determines the characteristic element, a criterion c which allows constructing subsets upon a set and the height of the output tree. To determine the characteristic element of a set, function f computes the arithmetic mean of all $p \in S$. In other words, f calculates coordinates of the centroid of all pixels in S . Each tree node is thus characterized by appropriate centroid. A criterion c allowing to create two subsets of S is defined as follows. Let $\rho(o_1, o_2)$ be Euclidean distance between points o_1 and o_2 . Let $m = \text{median}(\{\rho(p, f(S)): p \in S\})$. Put another way, m stands for a median distance between pixels in S and its centroid. Then, $c(S) = (S_1, S_2)$, where $S_1 = \{p \in S: \rho(p, f(S)) \leq m\}$ and $S_2 = \{p \in S: \rho(p, f(S)) \geq m\}$. In this way, S_1 contains pixels closer to the centroid (or equally distant) than m and S_2 contains pixels farther from the centroid (or equally distant) than m . Taking advantage of median ensures similar cardinality of subsets. As there are two subsets created in a step of our procedure, output tree T is a perfect binary one. The algorithm successively applies function f to determine characteristic elements for consecutive sets extracted with use of criterion c .

The tree created using the above algorithm describes the content of the images in such a way that for unlike images it produces unlike trees while trees of similar

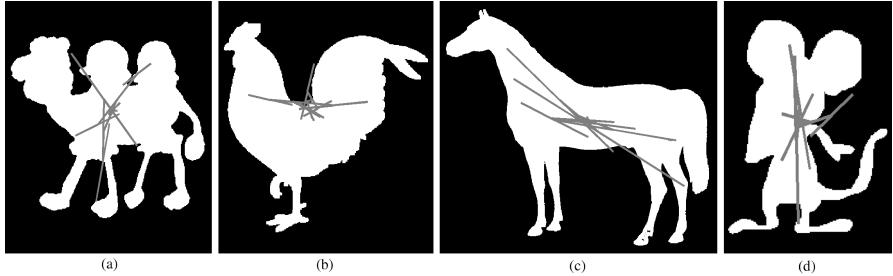


Fig. 1 Different binary images imply different trees. Test images of MPEG7 database [1]: (a) camel, (b) chicken, (c) horse and (d) rat exemplify the statement. The trees are of height $h = 4$.

images are also similar one to another. Examples of trees created upon four various binary shapes are shown in Fig. 1.

Following the main idea of the proposed approach, in order to find the dissimilarity of two binary images one computes the dissimilarity of trees obtained for images being compared. Since the way of creating the tree effects strict structure of the tree, while investigating differences between two trees, nodes can be compared pairwise. This is due to the fact that every node has its corresponding one in another tree. Difference between the trees is computed as a sum of differences between pairs of nodes. This way, the tree dissimilarity is expressed as a single scalar value.

Let T_1 and T_2 be trees of height h_1 and h_2 , respectively. The difference between the trees can be computed as:

$$D = \sum_{i=1}^l \delta_i \quad (1)$$

where δ_i stands for the difference of vertex pair i , $l = 2^{h+1} - 1$ with $h \in \mathbb{N}_+$ and $h \leq \min(h_1, h_2)$. The difference δ_i can be viewed as a “work” needed to translate a centroid of a node i to the position of centroid of corresponding node of another tree and vice versa. Term “work” is an analogy to the physical quantity, which is directly proportional to the force and the Euclidean distance between the centroids of a pair of corresponding vertices. The force is proportional to the number of pixels z_{ei} upon which the centroid for a vertex i in tree T_e is calculated.

Moreover, we take advantage of information about pixel distribution. For the centroid i , the information comprises average pixel distance \hat{a}_{ei} from the centroid i and standard deviation \hat{s}_{ei} of pixel distance from \hat{a}_{ei} . This reasoning leads to the formula:

$$\delta_i = \left(\frac{z_{1i}}{Z_1} + \frac{z_{2i}}{Z_2} \right) \left(d_i + \hat{a}_i + \hat{s}_i \right) \quad (2)$$

where $Z_e = \sum_{i=1}^l z_{ei}$, d stands for the distance, $\hat{a}_i = |\hat{a}_{1i} - \hat{a}_{2i}|$ and $\hat{s}_i = |\hat{s}_{1i} - \hat{s}_{2i}|$.

The above approach may be used directly only if the trees represent images of the same orientation and size. In case of images transformed with use of an affine transformation, like rotation or scaling, the trees must be matched prior to the computation of dissimilarity. In order to do this, first of all, trees are aligned, so that their roots are positioned in the origin of the coordinate system. It is achieved by translating all the vertices of a tree by the same vector applied to translate the root of the tree to the origin of the coordinate system. This ensures translation invariance of the trees. Then, tree T_1 is rotated and scaled. This is achieved by transforming coordinates of all vertices of the tree to the polar coordinate system, in which each point is determined by an angle from a fixed direction and a distance from the pole. For every vertex of the tree, the angle is increased by α and the distance is multiplied by s . This way, a new tree T_1 arises. Next, the difference D between T_1 and T_2 is calculated. It can be proven that invariance to scaling can be achieved by modifying the formula 2 so it takes the following form:

$$\delta_i = \left(\frac{z_{1i}}{Z_1^{1.5}} \frac{1}{s} + \frac{z_{2i}}{Z_2^{1.5}} \right) (d_i + \hat{a}_i + \hat{s}_i). \quad (3)$$

The goal is to find such an angle α and a scale s , that the difference D is minimal. Then, D specifies the dissimilarity of trees, thus the dissimilarity of images. When comparing trees created upon the same images (even if rotated or scaled), D equals 0 (or almost 0 due to rotation and scaling approximation). The difference between the trees increases along with the growth of the difference between the images and is—theoretically—unlimited. Therefore, $D \in \langle 0, \infty \rangle$.

3 Optimization Algorithms

Determining angle α and scale s in (3) is an optimization problem, that is the actual issue discussed in this paper. The optimized function is nonlinear, can have multiple minima, and does not possess derivatives. For such functions, heuristic optimization algorithms are usually the best choice. In this research, we take advantage of three heuristic optimization algorithms. Namely, genetic algorithm, particle swarm optimization, and simulated annealing.

Genetic algorithm (GA) [6], [5], [11] is a classic heuristic optimization approach. It follows the concept of population advancement based on natural evolution. A population composed of individuals (called also chromosomes) evolves to find the best argument value of the objective function. Individuals undergo mechanisms derived from evolution, in order to lead the population to better results. The algorithm works as follows. At first, population is randomly initiated in the search space. Then, in each iteration, chromosomes, every of which encodes a solution, are subjected to selection, gene recombination and mutation. These imitate evolution of a real, biological species. In effect, next generation replaces the previous one. Selection determines pairs of parent chromosomes, upon which children arise by means of gene crossover. Generally, the better is the value of the fitness function for a particular

individual, the greater are chances for it to become a parent. Crossover operator defines how genes of parent individuals are recombined. Mutation decreases the risk of the population getting stuck in a local minimum. The main parameters are population size P , maximal number of generations G and fraction of the population which undergoes crossover f_1 and mutation f_2 .

Particle swarm optimization (PSO) [8], [13] is a well-known optimization algorithm. It is an evolutionary computation method, which imitates social behaviour of a swarm of creatures. The idea benefits from advancement of a population of particles, which rove in the search space. Position of each particle represents a solution. Particles adjust their movement parameters according to their own experience as well as the experience of other particles of the swarm. A basic variant of PSO works as follows. First of all, population of particles is initiated. Particles, denoted below as x , are randomly positioned over a search space and are assigned random velocities. At iteration $k+1$ position of a particle i is updated according to the formula:

$$x_{k+1}^i = x_k^i + v_{k+1}^i \quad (4)$$

with the velocity v_{k+1}^i calculated as follows:

$$v_{k+1}^i = w_k v_k^i + c_{1k} r_1 (p_k^i - x_k^i) + c_{2k} r_2 (p_k^g - x_k^i) \quad (5)$$

where w_k stands for inertia weight, c_{1k} and c_{2k} stand for cognitive attraction and social attraction parameter, respectively. p_k^i is the best solution particle i ever found, p_k^g is the best solution ever found by any particle of the swarm, r_1 and r_2 are random numbers uniformly distributed on the interval $(0, 1)$. Inertia weight w_k as well as attraction parameters c_{1k} and c_{2k} can be constant or may be changed in the course of the run. The algorithm works and improves the solution until stopping criterion is met e.g. satisfactory solution achieved, inconsiderable solution improvement or number of iterations performed. PSO has gained an exceptional interest. There are many works introducing modifications to increase the performance of PSO and providing analysis of the algorithm and e.g. [4], [3]. Most of them relate to updating particle velocity.

Simulated annealing (SA) [9], [2] is another, heuristic global optimization approach. It is inspired by the process of annealing in metallurgy, where heating and cooling down the alloy is performed, in order to achieve the best structure of the material, therefore minimizing its energy. Heating shoves the molecules from their initial positions. This prevents getting stuck in a local minimum. On the other hand, cooling let them go into the state of lower energy than before, which means they head towards optimal solution. In each step of the algorithm, a new point is randomly generated. Its distance to the current point is based on a probability distribution which scale is proportional to the temperature. The current solution is replaced by a new one, when the latter lowers the objective (cooling down the material). However, with a certain probability, a new point may be accepted, even if it raises the objective (heating). The temperature is decreased in the run of the algorithm. This way, the extent of search is reduced in the final stages. The main parameters

are the initial temperature and the cooling function, which determines the way of decreasing the temperature.

4 Experimental Results

We performed experiments to juxtapose the three optimization approaches, for different parameter adjustments. We created a dataset of binary images¹, basing on MPEG7 shape database [1]. Our dataset is composed of 50 classes of images. Every class contains an original object as well as its variants, obtained by binary morphological operators of opening and closing [12]. Therefore, there are 150 images in the dataset. Every morphologically modified image has been scaled by a random factor in range $\langle 0.5, 2 \rangle$ and rotated by a random factor in range $\langle 0, 2\pi \rangle$. Upon the images, trees of height $h = 3$ were created. Evaluation procedure ran as follows. For a given optimization algorithm, for its certain configuration, every possible pair of trees was compared. As a result, we obtained square matrix of dissimilarity values. If an effective optimization method was used, then the values calculated for objects of the same class were small, whereas those calculated for objects from among different classes were greater. To reason about the efficiency of considered algorithm, we subjected the matrix of dissimilarity values to cluster analysis. The experiments were performed in Matlab environment.

Clustering was performed with use of k -medoids algorithm [7]. The method—on the basis of dissimilarity matrix—partitions the set of observations into predefined number of clusters. It is similar to k -means algorithm [14], [10]. However, in contrast to k -means, k -medoids selects cluster centers from among the observations rather than calculates centroids. The chosen centers (called medoids) define clusters in such a way that each observation is assigned to the least dissimilar medoid. The objective of the algorithm is to find such a configuration of medoids for which the sum of dissimilarities between the observations and their centers is minimal.

We express the reliability as the percentage of correctly clustered images. It may be difficult, however, to determine the percentage. Clustering algorithms deliver data grouped into clusters, but without labels. With different observations within a single cluster, it is tricky to assign the labels. For every case, we aimed to find such a label arrangement, that the percentage was highest.

The comparative study of the algorithms follows the reasoning, that the better an optimization method performed, the higher was the outcome percentage result. The successive tables present results obtained for considered algorithms, for different parameter configuration. Parameters were arranged in such a way, that maximal number of function evaluation was fixed. In effect, results of comparable computational effort for every method were achieved.

In case of genetic algorithm, each chromosome encodes a scale s and an angle α . The population evolves in search for the optimal values of these parameters. GA has

¹ Available at <http://www.isep.pw.edu.pl/~zielinsb/binary.zip>

been applied with different values of population size P , maximal number of generations G and crossover fraction f_1 . Mutation fraction f_2 equalled 0.01. Crossover was done by calculating weighted average of parents. Mutation was realized in an adaptive way. Table 1 presents the results.

Table 1 Percentage Results of GA for Different Parameters—Population Size, Generations, Crossover Fraction f_1 and Mutation Fraction $f_2 = 0.01$

Population	Generations								
	10			15			20		
	f_1			f_1			f_1		
	0.7	0.8	0.9	0.7	0.8	0.9	0.7	0.8	0.9
5	55.33	58.00	43.33	63.33	61.33	42.00	61.33	63.33	46.67
10	71.33	71.33	60.00	76.00	75.33	66.00	76.00	78.67	74.00
15	82.67	78.67	62.67	84.00	84.00	65.33	87.33	84.67	68.00

In the particle swarm optimization case we assumed that each particle's position represents a scale s and an angle α . Swarm proceeds to find the optimal configuration of the two parameters. We have applied the PSO algorithm with different values of population size P , maximal number of generations K , cognitive attraction c_{1k} and social attraction c_{2k} . Initial inertia weight $w_1 = 0.9$ and $w_k = w_1 - 0.4(k-1)/(K-1)$, which means that w_k decreased linearly from 0.9 at the first iteration to 0.5 at iteration K . PSO was made to stop if there was no improvement in the objective function for 5 consecutive generations. Tables 2, 3 and 4 present the results.

Table 2 Percentage Results of PSO for Different Parameters—Population Size, Generations, Cognitive Attraction c_{1k} and Social Attraction $c_{2k} = 0.75$

Population	Generations								
	10			15			20		
	c_{1k}			c_{1k}			c_{1k}		
	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
5	90.67	86.67	94.67	86.00	87.33	92.00	92.00	94.67	90.00
10	94.00	91.33	94.00	96.00	98.00	94.00	99.33	94.67	96.67
15	96.67	96.67	97.33	94.67	100.00	98.67	96.67	97.33	100.00

Table 3 Percentage Results of PSO for Different Parameters—Population Size, Generations, Cognitive Attraction c_{1k} and Social Attraction $c_{2k} = 1.25$

Population	Generations								
	10			15			20		
	c_{1k}			c_{1k}			c_{1k}		
	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
5	88.00	90.00	88.67	91.33	94.67	95.33	95.33	94.67	93.33
10	95.33	96.67	96.67	99.33	94.67	97.33	96.00	99.33	100.00
15	94.67	96.00	96.67	100.00	96.67	100.00	97.33	99.33	100.00

Table 4 Percentage Results of PSO for Different Parameters—Population Size, Generations, Cognitive Attraction c_{1k} and Social Attraction $c_{2k} = 1.75$

Population	Generations								
	10			15			20		
	c_{1k}			c_{1k}			c_{1k}		
	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
5	87.33	86.00	90.67	89.33	92.00	90.00	92.67	94.00	90.00
10	92.00	98.00	95.33	94.67	97.33	98.67	100.00	100.00	99.33
15	97.33	98.67	96.67	94.67	100.00	96.67	100.00	96.67	100.00

As far as simulated annealing is concerned, we optimized our black-box function with respect to scale s and angle α . The maximal number of objective function evaluations in a single run was fixed. This way, the algorithm was able to calculate the value of the function as many times as the tested population based methods. The initial temperature was set to either $t_1 = 50$ or $t_1 = 100$ and its value at iteration k was calculated as $t_k = 0.95^k * t_1$. Table 5 shows the results for both t_1 values and different numbers of objective function evaluations.

Experiments have been performed for small population sizes and numbers of generations, for both population based methods. This limited the number of possible evaluations of the objective function. For the simulated annealing approach, the number of the objective function evaluations was also limited. In effect, every approach was able to calculate the objective function at most the same number of times. This way, results are comparable. The experiments show, that the particle swarm optimization achieved the best results, which were satisfactory for small population sizes and numbers of generations. Simulated annealing was slightly worse. Both methods significantly outperformed the genetic algorithm.

Table 5 Percentage Results of SA Approach for Different Initial Temperature Values and Numbers of Objective Function Evaluations N

Initial Temperature	N						
	50	75	100	150	200	225	300
50	82.67	83.33	90.67	91.33	92.00	89.33	97.33
100	78.00	88.00	86.67	86.67	94.00	94.67	97.33

It is clearly visible, that the genetic algorithm achieved better results for greater population size and number of generations. Crossover fraction also played a significant role in the performance of the method. As for values $f_1 = 0.7$ and $f_1 = 0.8$ the results have been similar, increasing it to $f_1 = 0.9$ has led to the degradation of the results.

In the PSO evaluation, the size of the population has had a positive impact on the outcome. On the other hand, for the number of generations, social and cognitive attraction, we have observed no obvious influence of the parameters on the final outcome.

As one might expected, for the simulated annealing approach, generally, greater number of possible objective function evaluations has yielded better optimization results.

5 Conclusions

The aim of this research was to find out how heuristic optimization algorithms perform on a black-box function, for comparing binary images, with use of tree-based binary dissimilarity measure. We have evaluated the genetic algorithm, the particle swarm optimization and the simulated annealing approach on a function which allows determining a difference between trees. The tests have been applied for a different parameter values of considered methods. The results show the advantage of the particle swarm optimization and the simulated annealing algorithms over the genetic algorithm. Nevertheless, the best results have been obtained for the PSO approach.

References

1. MPEG7 CE Shape-1 Part B,
[http://imageprocessingplace.com/downloads_V3/
root_downloads/image_databases/MPEG7_CE-Shape-1_Part_B.zip](http://imageprocessingplace.com/downloads_V3/root_downloads/image_databases/MPEG7_CE-Shape-1_Part_B.zip)
(accessed in January 2013)
2. Cerny, V.: Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. Journal of Optimization Theory and Applications (1985)

3. Clerc, M., Kennedy, J.: The particle swarm—explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation* (2002)
4. Eberhart, R., Shi, Y.: Comparing inertia weights and constriction factors in particle swarm optimization. In: *Proceedings of the, Congress on Evolutionary Computation* (2000)
5. Goldberg, D.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston (1989)
6. Holland, J.: *Adaptation in Natural and Artificial Systems*. University of Michigan Press (1975)
7. Kaufman, L., Rousseeuw, P.: *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley–Interscience, New York (1990)
8. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: *Proceedings of IEEE International Conference on Neural Networks* (1995)
9. Kirkpatrick, S., Gelatt, C., Vecchi, M.: Optimization by simulated annealing. *Science* (1983)
10. MacQueen, J.: Some Methods for Classification and Analysis of Multivariate Observations. In: *Fifth Berkeley Symposium on Mathematical Statistics and Probability* (1967)
11. Michalewicz, Z.: *Genetic algorithms + data structures = evolution programs*. Springer, London (1996)
12. Serra, J.: *Image Analysis and Mathematical Morphology*. Academic Press, London (1982)
13. Shi, Y., Eberhart, R.: Empirical study of particle swarm optimization. In: *Proceedings of the 1999 Congress on Evolutionary Computation* (1999)
14. Steinhäus, H.: Sur la Division des Corps Materiél en Parties. *Bull. Acad. Polon. Sci., C1. III* (1956)
15. Zieliński, B., Iwanowski, M.: *Binary Image Comparison with Use of Tree-Based Approach*. Springer, Heidelberg (2013)
16. Zieliński, B., Iwanowski, M.: *Comparing Image Objects Using Tree-Based Approach*. Springer, Heidelberg (2012)

Model of Syntactic Recognition of Distorted String Patterns with the Help of GDPLL(k)-Based Automata

Janusz Jurek and Tomasz Peszek

Abstract. The process of syntactic pattern recognition consists of two main phases. In the first one the symbolic representation of a pattern is created (so called primitives are identified). In the second phase the representation is analyzed by a formal automaton on the base of a previously defined formal grammar (i.e. syntax analysis / parsing is performed). One of the main problems of syntactic pattern recognition is the analysis of distorted (fuzzy) patterns. If a pattern is distorted and the results of the first phase are wrong, then the second phase usually will not bring satisfactory results either. In this paper we present a model that could allow to solve the problem by involving an uncertainty factor (fuzziness/distortion) into the whole process of syntactic pattern recognition. The model is a hybrid one (based on artificial neural networks and GDPLL(k)-based automata) and it covers both phases of the recognition process (primitives' identification and syntax analysis). We discuss the application area of this model, as well as the goals of further research.

1 Introduction

Methods of pattern recognition and analysis could be divided into two classes [7, 18]:

1. Classical decision-theoretic methods. A pattern is treated in whole. The particular features of a pattern are extracted and analyzed. The recognition is based on formerly collected knowledge about the features.
2. Structural/syntactic methods. A pattern is treated as a structure consisting of so-called primitives. The recognition is based on knowledge about the possible structures.

Janusz Jurek · Tomasz Peszek

Information Technology Systems Department, Jagiellonian University
Cracow 30-348, ul. prof. St. Lojasiewicza 4, Poland

In a classical approach, a pattern is usually represented as a feature vector, which is subsequently analyzed (with regards to distinguished features) by means of various mathematical methods. In the structural/syntactic approach, pattern recognition consists of two main phases. In the first one the symbolic representation of a pattern is created: the basic components of the pattern (primitives) and relations between them are detected (in a classical way). In the second phase the structural representation is analyzed and recognized. Syntactic approach assumes the use of formal grammars and automata (parsers) in this phase. The applications of structural/syntactic methods include: chromosome analysis, bubble chamber tracks recognition, analysis of contours in 2D pictures, ECG and EEG analysis, speech recognition and many others [2, 7, 12, 13, 19].

One of the main practical problems with the applications of the syntactic pattern recognition methods concerns distortions (noise) in the analyzed patterns. If a pattern is distorted and the results of the first phase are wrong or inaccurate, then the second phase usually will not bring satisfactory results either. Let us stress, that even one misrecognized primitive could lead to the failure of the syntax analysis of the whole pattern. There are several methods of solving the problem known in the literature (e.g. [7]), but none of them satisfies *all* the following conditions:

- the discriminative (generating) power of the proposed grammars is bigger than context-free grammars,
- the syntax analyzer (parser) for the grammars is efficient computationally,
- the uncertainty factor (fuzziness/distortion) is taken into account in the whole process of recognition, i.e. both in the phase of primitives identification and the phase of the syntax analysis.

First two conditions are really important if we consider practical applications [6]. On the other hand, they are very difficult to fulfil since they are contradictory (i.e. the bigger descriptive power of a grammar the bigger computational complexity of a corresponding syntax analyzer is expected). Our approach to this problem is based on our previous research in the field of quasi-context sensitive string grammars [4,9].

As far as the third condition is concerned, we present a hybrid model that is based on artificial neural networks [17] and GDPLL(k) automata [5]. In the first phase of recognition we use probabilistic neural networks [16]. This tool can provide not only the identification of the primitives but also the probability distribution that a primitive belongs to a certain class. In the second phase we use a extended version of GDPLL(k) automaton that allow to analyze the whole pattern taking advantage of the information on the uncertainty of identified primitives (delivered by a probabilistic neural network). We equip a GDPLL(k) automaton with new, additional element, objective function, which could direct the extended analysis process.

Let us notice that our research into construction a described above model has been caused by the difficult constraints of the application of syntactic pattern recognition methods for the short-term electrical load forecasting (STLF) [11, 14]. The research into providing tools for such an application has been led by the authors for several years. However, it worth to point out that the application area of the model could be possibly very wide.

Basic definitions are presented in section 2. Section 3 contains the description of the architecture of the proposed model. Possible application areas are discussed in section 4. Final remarks are included in section 5.

2 Basic Definitions

In this section we briefly describe the context of the research and basic definition needed to present results achieved.

2.1 Probabilistic Neural Networks

The concept of probabilistic neural networks (PNN) was introduced by Specht in [16] in the beginning of 1990. Apart from classification, the networks could deliver some additional information on how far the presented pattern is related to all classes (not only to the winner one). This type of network consist of following layers.

1. *Input layer* where the pattern is sent in order to present it to the network.
2. *Hidden neurons layer* representing the population of patterns that have been already presented to the network (it counts the distances between currently presented pattern and the former ones. Each neuron in this layer connected to exactly one node in the subsequent layer (this implements the assignment of pattern to the class)).
3. *Classification layer* which is built of neurons corresponding to the classes (created on base of patterns assigned to the subsequent nodes in this layer). The activation of neurons in this layer is counted on the base of the number and the activation power of nodes from the previous layer connected to the current node. The neurons that have the greatest power of activation determine the classes, to which the presented pattern with highest probability belongs. We can see the activation of the network as the probability distribution of random variable, which describes the class to which the presented patterns belongs.
4. *Optional decision rule layer* (node), by means of which the decision rule is implemented. It receives the information on the activation of the previous layer.

Typical applications of PNN are described in [3] or in [8] (where a similar architecture is presented). The application area is very wide and it includes: voice identification, face recognition, signal processing, tissue diagnostics in medicine, texture classification, cloud classification and many others.

The main advantages of PNN (when comparing to classical NN) are exactness and rapid answer, relative small sensitivity to outliers, tendency to become the optimal Bayes classifier (with growth of learning set). As main weaknesses of PNN one can indicate slower classification of new patterns and higher storage requirements (the number of neurons comparable with the quantity of presented patterns).

2.2 *GDPLL(k) Grammars and Automata*

Let us introduce two basic definitions corresponding to GDPLL(k) grammars [4, 9].

Definition 1. A *generalized dynamically programmed context-free grammar* is a six-tuple $G = (V, \Sigma, O, P, S, M)$, where: V is a finite, nonempty alphabet; $\Sigma \subset V$ is a finite, nonempty set of terminal symbols (let $N = V \setminus \Sigma$); O is a set of basic operations on the values stored in the memory; $S \in N$ is the starting symbol; M is the memory; P is a finite set of productions of the form: $p_i = (\mu_i, L_i, R_i, A_i)$ in which $\mu_i : M \rightarrow \{\text{TRUE}, \text{FALSE}\}$ is the predicate of applicability of the production p_i defined with the use of operations ($\in O$) performed over M ; $L_i \in N$ and $R_i \in V^*$ are left- and right-hand sides of p_i respectively; A_i is the sequence of operations ($\in O$) over M , which should be performed if the production is to be applied. \square

Definition 2. Let $G = (V, \Sigma, O, P, S, M)$ be a dynamically programmed context-free grammar. The grammar G is called a *GDPLL(k) grammar*, if the following two conditions are fulfilled.

1. Stearns's condition of LL(k) grammars. (The top-down left-hand side derivation is deterministic if it is allowed to look at k input symbols to the right of the current position of the input head in the string).
2. There exists a certain number ξ such that after the application of ξ productions in a left-hand side derivation we get at the "left-hand side" of a sentence at least one new terminal symbol. \square

A derivation in GDPLL(k) grammars is defined in the following way. Before application of a production p_i we test whether L_i occurs in a sentential form derived. Then we check the predicate of applicability of the production. The predicate is defined as an expression based on variables stored in the memory. If the predicate is true, we replace L_i with R_i and then we perform the sequence of operations over the memory. The execution of the operations changes the contents of the memory. It is done with the help of arithmetical and assignment instructions.

The automaton for GDPLL(k) grammars (GDPLL(k) parser) has been described in [9]. Its algorithm reflects the way how the derivation in the grammars is performed (top-down approach is used).

GDPLL(k) grammars are characterized by three main features: they are of very good discriminative properties (they are able to generate a large class of context-sensitive languages) [4], languages generated by them can be efficiently analyzed i.e. it is possible to construct an automaton of the linear computational complexity [9], and there is a grammatical inference algorithm for GDPLL(k) grammars [10] (a feature also very important in practical applications [6]). Finally, let us point out that a GDPLL(k) grammar could be designed as stochastic one [5].

3 Hybrid Model of Syntactic Recognition of Distorted Patterns

In this section we present a pattern recognition model that could be a solution to the problem described in the introduction, i.e. the problem of distortions (fuzziness/noise) in analyzed patterns. The model is a hybrid one: it is based on probabilistic neural networks and GDPLL(k)-based automata with an additional objective function. The model covers two phases of the recognition process: primitives' identification and syntax analysis.

3.1 Primitive Identification

A probabilistic neural network (PNN) should deliver the symbolic representation of a pattern. In a classical syntactic recognition approach this representation is in the form of a word (string) built of the symbols from a given alphabet. A PNN in our model delivers additionally the distribution describing the probability that the recognized symbol belongs to given classes.

So instead of a word of alphabet letters, the word of vectors is generated. The vectors are of the form: $((a_1, p_1), \dots, (a_n, p_n))$, where a_i is the possible symbolic value of the primitive, and p_i the probability that the primitive is equal a_i .

The syntax analyzer gets as an input:

- a whole word consisting of such vectors (in case of an off-line system designed for recognizing whole patterns) or
- a group of vectors determining a fragment of the word (in case of an online system, where the immediate recognition of a fragment of a pattern is needed).

3.2 Syntax Analysis

A word of vectors delivered by PNN can be resolved by many different strings, so potentially large number of strings will require syntax analysis. The problem is how to analyze such a big set, considering a subset large enough not to skip important information and not to lose analysis efficiency.

An analysis should be conducted by specially designed parser. Its idea is based on a GDPLL(k) parser described in [4]. The parser architecture is depicted in figure 1. It consists of the following main parts:

- The **input tape** containing a word of vectors. Every cell contains a single vector recognized by PNN.
- The **head** reading the current cell of the input tape (or k -subsequent cells) and forwarding the information to the control module.
- The collection of **GDPLL(k) parsers** performing syntax analysis on base of theirs (the same) control table.

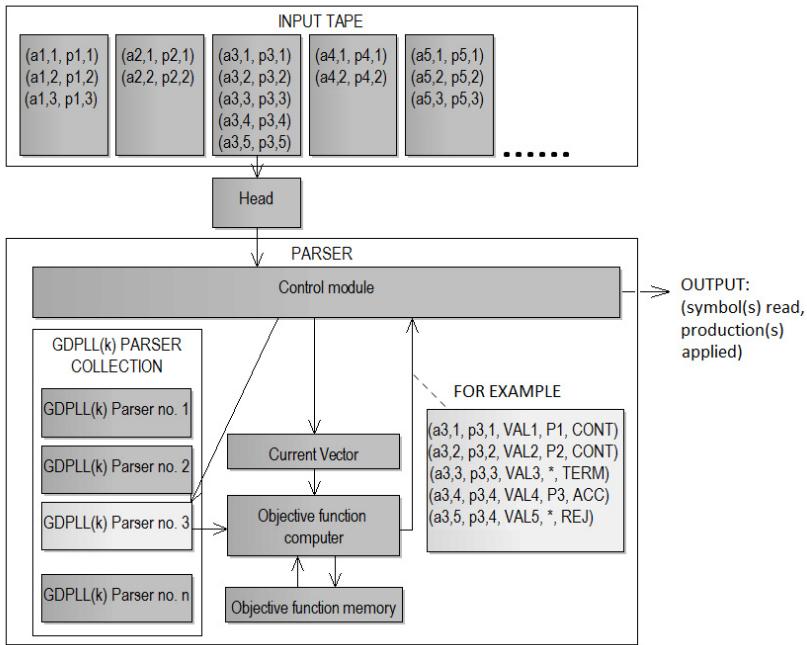


Fig. 1 The architecture of the parser

- The **objective function computer**: a subautomaton computing the objective function value for the current GDPLL(k) parser, and the current input vector. The output of the subautomaton is of the following form: $(a_1, p_1, val_1, prod_1, action_1), \dots, (a_n, p_n, val_n, prod_n, action_n)$, where: (a_i, p_i) is a component of an input vector, val_i is the objective function value, $prod_i$ is the production to be used in the derivation process, and $action_i$ is one of the control actions:
 - **ACC**: a_i is successfully derived and the whole input word is recognized (accepted),
 - **REJ**: it is impossible to derive a_i , so the derivation path should be rejected,
 - **TERM**: it is possible to derive a_i , but the derivation path should be terminated because the objective function value is not high enough,
 - **CONT**: it is possible to derive a_i , starting with the application of the production $prod_i$, add the derivation along the path should be continued (the objective function value is high).
- The **memory** of the objective function computer which is needed for auxiliary data collected while computing function values.
- The **control module** managing the whole analysis process. As the automaton conducts several derivations parallel, the control module has to manage the collection of GDPLL(k) parsers.

- The **output of the control module** delivering a user information on symbol (or symbols) read and production (or productions) applied to derive these symbols.

The control module works in the following iterations:

- 1 It reads the subsequent vector (or vectors) from the input tape to the memory,
- 2 For each active GDPLL(k) parser in the collection (ordered by computed analysis priority) it verifies the possible paths of further analysis:
- 2a For each symbol from the input vector (ordered by recognition probability) it finds the productions that can lead to derive this symbol (with the use of a GDPLL(k) parser). The pair: (*current symbol, production deriving this symbol*) uniquely determines the possible path of further analysis.
- 2b The value of objective function is computed and stored for the pair (with the use of the objective function computer). Each pair is supplied by an additional information: an objective function value, a production (or production list) needed to derive particular symbol and a further action (i.e. *ACC, REJ, CONT, TERM*).
- 3 For determined possible paths of further analysis the control module performs:
 - a) If the action is *CONT* it changes status of the GDPLL(k) parsers or — if needed — creates new instances of GDPLL(k) parser. Newly created GDPLL(k) parsers are inserted into the parser collection. The collection is subsequently reordered on the base of computed objective function value. Let us notice, that the possible paths of further derivation may fork as a result of ambiguity in primitive recognition.
 - b) If the action is *ACC, REJ* or *TERM*, then the GDPLL(k) parser is terminated.
- 4 If all parsers in the collection are terminated or there is no more vectors to be read the parsing process is finished.

The objective function is the element that can direct the analysis process according to a desired strategy. Let us consider some examples of such strategies:

RecProb: We accept only the symbols with the highest recognition probability (from each vector delivered by PNN). It means that the value of objective function is positive only for such symbols.

LimCost: This strategy allows the analysis of the symbols with the lower recognition probabilities. In such case, the higher recognition probability of the symbol, means the higher value of the objective function. However, we have to ensure that the parsing remains efficient even if we allow many paths of the derivation. One of the solutions to this problem is the following method. The control module maintains the maximal number of active GDPLL(k) parsers. Periodically (according to an established parameter) the parsers collection is reviewed by the control module and the most promising parser (derivation path) is chosen (according to the cumulative value of the objective function). The other parsers (derivation paths) are terminated and the analysis is carried on only by the one chosen parser.

AddFact: In some applications it could be useful if we consider additional (expert) knowledge concerning the application during the derivation process. In this situation appropriate implementation of an objective function should help to include the extra information in the process of the analysis.

Let us notice that the first two strategies: *RecProb* and *LimCost* could guarantee that the good efficiency of the parsing process (resulting from the GDPLL(k) parser usage) will be preserved. As far as *AddFact* is concerned, the efficiency strictly depends on an application's constraints and the model of including the additional information into analysis.

The effect of the parser analysis is the recognition (acceptation or rejection) of the input word (in case of offline systems) or the information about its fragments (in case of online systems).

4 Possible Application Areas

As we mentioned in the introduction the model presented has been developed to satisfy difficult constraints of the application for the short-term electrical load forecasting (STLF) [1, 11, 14, 20]. STLF is very important for the effective management of power utilities. Exact prediction on expected load for the next hour, shift, day ahead allow flexible and effective allocation of resources needed for energy production (fuels, water, etc.). The application of syntactic pattern recognition methods to the analysis of the electrical load functions could be done in an analogical way, like in case of ECG or EEG analysis [7, 15]. We "translate" the functions to the symbolic form, and then we analyze the string of symbols obtained. An example of load functions are shown in Figure 2.

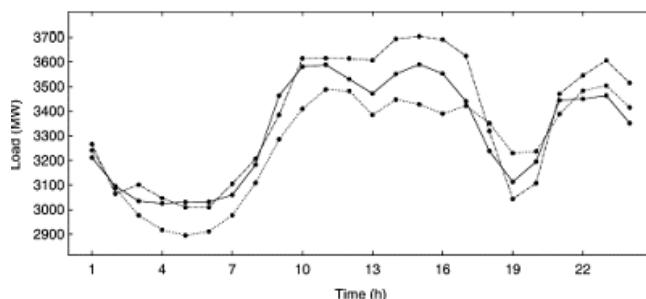


Fig. 2 Daily electrical load distribution

If the syntactic pattern recognition system for STLF is to be practically useful (especially in case of "non-typical" days like a day between holidays), it should have to deal with an uncertainty factor. Model presented could be a possible solution

to the problem, although implementation details and its practical verification still require further studies.

Although STLF has been the main inspiration for our research, the model could be applied in wide area of applications. If patterns to be recognized are distorted or noisy, then the model could be especially suitable as a base for building a syntactic pattern recognition system. What's more it could be adapted for the applications where on-line analysis is needed (thanks to the top-down analysis of an input word, generating "partial" recognitions).

It is also worth to notice that additionally the model delivers solutions for other main problems concerning the syntactic approach in pattern recognition, i.e. the problems of assuring big discriminative power and low computational complexity, which is very important from a practical point of view [6].

5 Concluding Remarks

In the paper we have presented a hybrid model of syntactic recognition of distorted string patterns. The advantages of the model are following: the grammars used are of big discriminative power, the syntax analyzer (parser) for the grammars is efficient computationally, and the uncertainty factor (caused by a possible distortion / fuzziness of a pattern) is taken into account both in the phase of primitives identification (done with the use of the probabilistic neural networks) and in the phase of the syntax analysis.

The syntax analyzer could be configured according to three different strategies: *RecProb*, *LimCost*, *AddFact*. The strategies can also be used for the online analysis of an "infinite" word. This makes it useful in the applications where we should recognize *fragments* of a pattern in a real time.

Further steps of our research are aimed at two goals. The first one is a definition of the algorithm of automatical inferencing stochastic GDPLL(k) grammars for distorted pattern analysis. The algorithm will be based on the model for GDPLL(k) grammars [10], but it will be enhanced to cover aspects of uncertainty. The second one is practical verification of the model in a short term electrical load forecast application. The results of the research will be the subject of further publications.

References

1. Alfares, H.K., Nazeeruddin, M.: Electric load forecasting: literature survey and classification of methods. *International Journal of Systems Science* 33, 23–34 (2002)
2. Bunke, H.O., Sanfeliu, A. (eds.): *Syntactic and Structural Pattern Recognition — Theory and Applications*. World Scientific, Singapore (1990)
3. Emary, I.M., Ramakrishnan, S.: On the Application of Various Probabilistic Neural Networks in Solving Different Pattern Classification Problems. *World Applied Sciences Journal* 4, 772–780 (2008)

4. Flasiński, M., Jurek, J.: Dynamically Programmed Automata for Quasi Context Sensitive Languages as a Tool for Inference Support in Pattern Recognition-Based Real-Time Control Expert Systems. *Pattern Recognition* 32, 671–690 (1999)
5. Flasiński, M., Jurek, J.: On the analysis of fuzzy string patterns with the help of extended and stochastic GDPLL(k) grammars. *Fundamenta Informaticae* 71, 1–14 (2006)
6. Flasiński, M., Jurek, J.: Syntactic Pattern Recognition: Survey of Frontiers and Crucial Methodological Issues. In: Burduk, R., Kurzyński, M., Woźniak, M., Żołnierk, A. (eds.) *Computer Recognition Systems 4. AISC*, vol. 95, pp. 187–196. Springer, Heidelberg (2011)
7. Fu, K.S.: *Syntactic Pattern Recognition and Applications*. Prentice Hall (1982)
8. Goodman, R.M., Higgins, C.M., Miller, J.W.: Rule-Based Neural Networks for Classification and Probability Estimation. *Neural Computation* 4, 781–804 (1992)
9. Jurek, J.: Recent developments of the syntactic pattern recognition model based on quasi-context sensitive languages. *Pattern Recognition Letters* 26, 1011–1018 (2005)
10. Jurek, J.: Grammatical Inference as a Tool for Constructing Self-learning Syntactic Pattern Recognition-Based Agents. In: Bubak, M., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) *ICCS 2008, Part III. LNCS*, vol. 5103, pp. 712–721. Springer, Heidelberg (2008)
11. Jurek, J., Peszek, T.: On the use of syntactic pattern recognition methods, neural networks, and fuzzy systems for short-term electrical load forecasting. *Advances in Soft Computing*, pp. 851–858. Springer, Heidelberg (2005)
12. Ogiela, M.R., Ogiela, U.: DNA-like linguistic secret sharing for strategic information systems. *International Journal of Information Management* 32, 175–181 (2012)
13. Pavlidis, T.: *Structural Pattern Recognition*. Springer (1977)
14. Peszek, T.: Neuro–Fuzzy Prediction Systems in Energetics. *Schedae Informaticae* 15, 73–94 (2006)
15. Pietka, E.: Feature extraction in computerized approach to the ecg analysis. *Pattern Recognition* 24, 139–146 (1991)
16. Specht, D.F.: Probabilistic Neural Networks. *Neural Networks* 3, 109–118 (1990)
17. Tadeusiewicz, R.: *Sieci neuronowe*. Akademicka Oficyna Wydawnicza, Warszawa (1993)
18. Tadeusiewicz, R., Flasiński, M.: *Rozpoznawanie Obrazów*. Państwowe Wydawnictwo Naukowe PWN, Warszawa (1991)
19. Tadeusiewicz, R., Ogiela, M.R.: *Medical Image Understanding Technology*. Springer, Heidelberg (2004)
20. Taylor, J., McSharry, P.: Short-Term Load Forecasting Methods: An Evaluation Based on European Data. *IEEE Transactions on Power Systems* 22, 2213–2219 (2008)

Decision Rules in Simple Decision Systems over Ontological Graphs

Krzysztof Pancerz

Abstract. In the paper, we consider decision rules in simple decision systems over ontological graphs. We assign ontological graphs extending the knowledge about attribute values (i.e., adding some semantic meanings of attribute values) to condition attributes in such systems. Semantics changes the view on properties of decision rules, especially, on the validity of rules in decision systems. Decision rules and related notions are defined analogously to those defined for classic decision systems in rough set theory.

1 Introduction

In the paper, we consider information systems (decision systems), from the Pawlak's perspective, as the knowledge representation systems (cf. [11]). In the classic approach, attribute values in information systems (decision systems) can be both categorical and continuous. Classic rough set theory is based on an indiscernibility relation defined over a set of objects, described by attribute value vectors, in a given information (decision) system (cf. [11]). Two objects are indiscernible if and only if for each attribute its value is the same for both objects. "The same" means that attribute values are expressed by the same symbol. Indiscernibility reflects a total impossibility of distinguishing between objects. In this case, we are interested in exact (literal, symbol-precise) meaning of attribute values. Such a look at attribute values can lead to some problems, especially, in case of categorical values on account of loss of their semantic meanings. To avoid nuisances in interpretation of meanings of attribute values, we propose to put attribute values into the ontological (semantic) space. Such extensions (improvements) have been considered in the

Krzysztof Pancerz

University of Management and Administration in Zamość, Poland

e-mail: kpancerz@wszia.edu.pl

University of Information Technology and Management in Rzeszów, Poland

literature, e.g., DAG-Decision Systems [5], Dominance-Based Rough Set Approach (DRSA) [2], Rough Ontology [3], etc. In our approach, we replace, in a classic definition of information (decision) systems, simple sets of attribute values by ontological graphs which deliver us some new knowledge about meanings of attribute values. For this case, decision rules in decision systems can be seen from different perspectives, for example, taking into consideration synonymy, generality or some more sophisticated properties determining meanings of attribute values. Replacing attribute value sets by ontological graphs leads to generalization of different notions defined in classic rough set theory. Therefore, notions defined in the presented paper are based on the classic ones.

2 Decision Systems over Ontological Graphs

We propose to consider attribute values in the ontological (semantic) space. Our approach is based on the definitions of ontology given by Neches et al. [7] and Kohler [4]. That is, ontology is constructed on the basis of a controlled vocabulary and the relationships of the concepts in the controlled vocabulary. Formally, the ontology can be represented by means of graph structures. In our approach, the graph representing the ontology \mathcal{O} is called the ontological graph. In such a graph, each node represents one concept from \mathcal{O} , whereas each edge represents a relation between two concepts from \mathcal{O} .

Definition 1. Let \mathcal{O} be a given ontology. An ontological graph is a quadruple $OG = (\mathcal{C}, E, \mathcal{R}, \rho)$, where \mathcal{C} is a nonempty, finite set of nodes representing concepts in the ontology \mathcal{O} , $E \subseteq \mathcal{C} \times \mathcal{C}$ is a finite set of edges representing relations between concepts from \mathcal{C} , \mathcal{R} is a family of semantic descriptions (in natural language) of types of relations (represented by edges) between concepts, and $\rho : E \rightarrow \mathcal{R}$ is a function assigning a semantic description of the relation to each edge.

Relations are very important components in ontology modeling as they describe the relationships that can be established between concepts. We touch the problem of semantic relations (meaning relations) considered in linguistics. In general, three basic kinds (families) of relations can be distinguished (cf. [6]): equivalence relations, hierarchical relations, associative relations. For example, synonymy is a kind of an equivalence relation. The genus-species ("is-a") relation is the traditional example of hierarchical relations. An associative relation is any non-hierarchical and non-equivalence relation. In the proposed approach, we take into consideration the following family of semantic descriptions of relations between concepts (cf. [9]): $\mathcal{R} = \{"\text{is synonymous with}", "\text{is generalized by}", "\text{is specified by}"\}$. Some other relations are considered in [9]. We will use the following notation: R_{\sim} - "is synonymous with", R_{\triangleleft} - "is generalized by", R_{\triangleright} - "is specified by". Therefore, for simplicity $\mathcal{R} = \{R_{\sim}, R_{\triangleleft}, R_{\triangleright}\}$.

According to categorization of relations between the concepts given above, we have that R_{\sim} belongs to the family of equivalence relations whereas R_{\triangleleft} and R_{\triangleright}

belong to the family of hierarchical relations. Moreover, relations have the following properties: R_{\sim} is reflexive, symmetric and transitive, R_{\triangleleft} is reflexive and transitive, R_{\triangleright} is reflexive and transitive. In the graphical representation of the ontological graph, for readability, we will omit reflexivity of relations. However, the above relations are reflexive, i.e., a given concept is synonymous with itself, a given concept is generalized by itself, a given concept is specified by itself.

We can create decision systems over the ontological graphs. It can be done in different ways. In our investigation, condition attribute values of a given decision system are concepts from ontologies assigned to attributes. Such a system is called a simple decision system over ontological graphs (cf. [9]). Below, we remind its definition.

Definition 2. A simple decision system SDS^{OG} over ontological graphs is a tuple $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$, where U is a nonempty, finite set of objects, C is a nonempty, finite set of condition attributes, D is a nonempty, finite set of decision attributes, $\{OG_a\}_{a \in C}$ is a family of ontological graphs associated with condition attributes from C , $V_d = \bigcup_{a \in D} V_a$, V_a is a set of values of the decision attribute $a \in D$, $c : C \times U \rightarrow \mathcal{C}$, $\mathcal{C} = \bigcup_{a \in C} \mathcal{C}_a$, is an information function such that $c(a, u) \in \mathcal{C}_a$ for each $a \in C$ and $u \in U$, \mathcal{C}_a is a set of concepts from the graph OG_a , $d : D \times U \rightarrow V_d$ is a decision function such that $d(a, u) \in V_d$ for each $a \in D$ and $u \in U$.

Remark 1. It is not necessary for an information function to be a total function, i.e., $c : U \times C \rightarrow \mathcal{C}^* \subseteq \mathcal{C}$.

Example 1. Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system, represented by a decision table (see Table 1), over ontological graphs shown in Figures 1 and 2. In this system, $U = \{u_1, u_2, \dots, u_8\}$ is a set of eight persons described with respect to their material status, $C = \{Vehicle, Dwelling, Job\}$ is a set of condition attributes describing selected possessions of persons, $D = \{Material Status\}$ is a set of decision attributes, D consists of one attribute evaluating material status of persons, $OG_{Vehicle} = (\mathcal{C}_V, E_V, \mathcal{R}, \rho_V)$ is an ontological graph associated with the attribute *Vehicle*, where

$$\mathcal{C}_V = \{Automobile, Bicycle, Car, Lorry, Minivan, SUV, Vehicle\}$$

and the rest of elements can be obtained from Figure 1 (a), $OG_{Dwelling} = (\mathcal{C}_D, E_D, \mathcal{R}, \rho_D)$ is an ontological graph associated with the attribute *Dwelling*, where

$$\mathcal{C}_D = \{Apartment, Bungalow, Condominium, Cottage, Dwelling, House, Flat, Logging\}$$

and the rest of elements can be obtained from Figure 1 (b), $OG_{Job} = (\mathcal{C}_J, E_J, \mathcal{R}, \rho_J)$ is an ontological graph associated with the attribute *Job*, where

$$\mathcal{C}_J = \{Academic, Clark, Mental worker, Office worker, Physical worker, Research worker, Worker\}$$

and the rest of elements can be obtained from Figure 2, $\mathcal{R} = \{R_\sim, R_\lhd, R_\rhd\}$, $V_d = \{\text{Low}, \text{Medium}, \text{High}\}$, c is an information function and d is a decision function, both defined in the tabular form in Table 1.

Table 1 A simple decision system over ontological graphs

$U/C \cup D$	<i>Vehicle</i>	<i>Dwelling</i>	<i>Job</i>	<i>Material Status</i>
u_1	<i>Car</i>	<i>Lodging</i>	<i>Physical worker</i>	<i>Medium</i>
u_2	<i>Minivan</i>	<i>House</i>	<i>Clerk</i>	<i>High</i>
u_3	<i>Car</i>	<i>Flat</i>	<i>Office worker</i>	<i>Medium</i>
u_4	<i>Bicycle</i>	<i>Lodging</i>	<i>Worker</i>	<i>Low</i>
u_5	<i>SUV</i>	<i>Bungalow</i>	<i>Academic</i>	<i>High</i>
u_6	<i>Car</i>	<i>Lodging</i>	<i>Physical worker</i>	<i>Low</i>
u_7	<i>Car</i>	<i>Condominium</i>	<i>Mental worker</i>	<i>Medium</i>
u_8	<i>Automobile</i>	<i>Cottage</i>	<i>Research worker</i>	<i>Medium</i>

Obviously, ontological graphs used in this example have been simplified in comparison to ontological graphs expressing real-world relations between concepts.

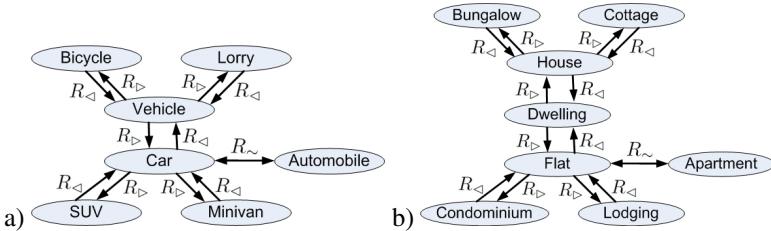


Fig. 1 (a) an ontological graph $OG_{Vehicle}$ associated with the attribute *Vehicle*, (b) an ontological graph $OG_{Dwelling}$ associated with the attribute *Dwelling*

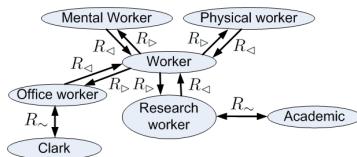


Fig. 2 An ontological graph OG_{Job} associated with the attribute *Job*

3 Relations over Sets of Attribute Values

In our approach, we propose to consider some relations defined over sets of attribute values in simple decision systems over ontological graphs. In defined relations, we use some additional knowledge about relationships between attribute values which is included in ontological graphs.

Definition 3. Let $OG_a = (\mathcal{C}_a, E_a, \mathcal{R}, \rho_a)$ be an ontological graph associated with the attribute a in a simple decision system, where $\mathcal{R} = \{R_\sim, R_\lhd, R_\rhd\}$. An exact (literal, symbol-precision) meaning relation between $c_1, c_2 \in \mathcal{C}_a$ is defined as $EMR(a) = \{(c_1, c_2) \in \mathcal{C}_a \times \mathcal{C}_a : c_1 = c_2\}$.

The exact meaning relation EMR is reflexive, symmetric and transitive (it is an equivalence relation).

Definition 4. Let $OG_a = (\mathcal{C}_a, E_a, \mathcal{R}, \rho_a)$ be an ontological graph associated with the attribute a in a simple decision system, where $\mathcal{R} = \{R_\sim, R_\lhd, R_\rhd\}$. A synonymous meaning relation between $c_1, c_2 \in \mathcal{C}_a$ is defined as $SMR(a) = \{(c_1, c_2) \in \mathcal{C}_a \times \mathcal{C}_a : (c_1, c_2) \in E_a \wedge \rho((c_1, c_2)) = R_\sim\}$.

The synonymous meaning relation SMR is reflexive, symmetric and transitive (it is an equivalence relation).

Let $OG = (\mathcal{C}, E, \mathcal{R}, \rho)$ be an ontological graph. In the next definitions, we will use the following notation: $[c_i, c_j]$ is a simple path in OG between $c_i, c_j \in \mathcal{C}$, $\mathcal{E}([c_i, c_j])$ is a set of all edges from E belonging to the simple path $[c_i, c_j]$, and $\mathcal{P}(OG)$ - is a set of all simple paths in OG .

Remark 2. In the literature, there are different definitions for a simple path in the graph. In this paper, we follow the definition in which a path is simple if no node or edge is repeated, with the possible exception that the first node is the same as the last. Therefore, the path $[c_i, c_j]$, where $c_i, c_j \in \mathcal{C}$ and $c_i = c_j$ can also be a simple path in OG .

Definition 5. Let $OG_a = (\mathcal{C}_a, E_a, \mathcal{R}, \rho_a)$ be an ontological graph associated with the attribute a in a simple decision system, where $\mathcal{R} = \{R_\sim, R_\lhd, R_\rhd\}$. A general meaning relation $GMR^k(a)$ of at most k -th order is a set of all pairs $(c_1, c_2) \in \mathcal{C}_a \times \mathcal{C}_a$ satisfying the following condition. There exists $c_3 \in \mathcal{C}_a$ such that the following holds:

$$\exists_{[c_1, c_3] \in \mathcal{P}(OG_a)} \left(\forall_{e \in \mathcal{E}([c_1, c_3])} \rho(e) \in \{R_\sim, R_\lhd\} \right)$$

and $(\text{card}(\{e' \in \mathcal{E}([c_1, c_3]) : \rho(e') = R_\rhd\}) \leq k)$,

and

$$\exists_{[c_2, c_3] \in \mathcal{P}(OG_a)} \left(\forall_{e \in \mathcal{E}([c_2, c_3])} \rho(e) \in \{R_\sim, R_\lhd\} \right)$$

and $(\text{card}(\{e' \in \mathcal{E}([c_2, c_3]) : \rho(e') = R_\rhd\}) \leq k)$.

The general meaning relation of at most k -th order is reflexive, symmetric and transitive (it is an equivalence relation). If c_1 is GMR^k -related to c_2 , then there exists $c_3 \in \mathcal{C}$ such that c_1 is generalized by c_3 and c_2 is generalized by c_3 , both through at most k concepts. In the generalization path, we also take into consideration synonyms, but they do not affect the order of generalization.

Example 2. Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system given in Example 1. Let us consider the attribute *Vehicle*. In our example, the following holds: $(Minivan, Car) \in GMR^1(Vehicle)$, because there exists $[Minivan, Car] \in \mathcal{P}(OG_{Vehicle})$ consisting of only one edge $(Minivan, Car)$ such that $\rho((Minivan, Car)) = R_\triangleleft$ and there exists $[Car, Car] \in \mathcal{P}(OG_{Vehicle})$ consisting of only one edge (Car, Car) such that $\rho((Car, Car)) = R_\triangleleft$ in view of reflexivity of R_\triangleleft .

Definition 6. Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system over ontological graphs, $a \in C$, R_a be a relation over \mathcal{C}_a , where \mathcal{C}_a is a set of concepts from the graph OG_a associated with a , and $c_1, c_2 \in \mathcal{C}_a$. A meaning equivalence \mathcal{M}_{R_a} of c_1 and c_2 with respect to R_a holds if and only if $(c_1, c_2) \in R_a$. This fact will be denoted by $\models \mathcal{M}_{R_a}(c_1, c_2)$.

We can distinguish the following meanings of condition attribute values:

- \mathcal{M}_\equiv - exact (literal, symbol-precise) meaning of attribute values. In the \mathcal{M}_\equiv meaning, attribute values are considered with respect to the exact meaning relation *EMR* (see Definition 3).
- \mathcal{M}_\sim - synonymous meaning of attribute values. In the \mathcal{M}_\sim meaning, attribute values are considered with respect to the synonymous meaning relation *SMR* (see Definition 4).
- $\mathcal{M}_\triangleleft^k$ - general meaning of at most k -th order. In the $\mathcal{M}_\triangleleft^k$ meaning, attribute values are considered with respect to the general meaning relation *GMR^k* of at most k -th order (see Definition 5).

Each meaning \mathcal{M} is associated with the corresponding relation R defined over the attribute value set. It will be denoted by \mathcal{M}_R .

Example 3. Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system given in Example 1. Let us consider the attribute *Vehicle*. A meaning equivalence $\mathcal{M}_\triangleleft^1$ of *Minivan* and *Car* holds because $(Minivan, Car) \in GMR^1(Vehicle)$, i.e., $\models \mathcal{M}_\triangleleft^1(Minivan, Car)$.

4 Decision Rules

In this section, we extend notions related to decision rules in classic decision systems, given in [11], to analogous notions for simple decision systems over ontological graphs. Another look, this time based on Dominance-Based Rough Set Approach

(DRSA) [2], at decision rules for simple decision systems over ontological graphs has been presented in [8].

Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system over ontological graphs. Let $\mathcal{C} = \bigcup_{a \in C} \mathcal{C}_a$, where \mathcal{C}_a is a set of concepts from the graph OG_a

associated with a given $a \in C$. For the decision system SDS^{OG} , we define condition descriptors which are expressions (a, v) over C and \mathcal{C} , where $a \in C$ and $v \in \mathcal{C}$ as well as decision descriptors which are expressions (a, v) over D and V_d , where $a \in D$ and $v \in V_d$. The set $\mathcal{F}(C, \mathcal{C})$ of formulae over C and \mathcal{C} is the least set containing all atomic formulae over C and \mathcal{C} and closed with respect to the propositional connectives \wedge (conjunction), \vee (disjunction) and \neg (negation). Analogously, the set $\mathcal{F}(D, V_d)$ of formulae over D and V_d is the least set containing all atomic formulae over D and V_d and closed with respect to the propositional connectives \wedge , \vee and \neg . The formulae from $\mathcal{F}(C, \mathcal{C})$ and $\mathcal{F}(D, V_d)$ are called condition formulae and decision formulae of SDS^{OG} , respectively.

Definition 7. Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system over ontological graphs. A condition attribute value meaning map associated with the set C of condition attributes from SDS^{OG} is an ordered tuple $\mathcal{M}(C) = (\mathcal{M}_{R_1}, \mathcal{M}_{R_2}, \dots, \mathcal{M}_{R_l})$, where $l = \text{card}(C)$, describing assignment of meaning relations to condition attributes from SDS^{OG} , i.e., \mathcal{M}_{R_i} is assigned to $a_i \in C$, where $i = 1, 2, \dots, l$.

Definition 8. Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system over ontological graphs, $\mathcal{F}(C, \mathcal{C})$ be the set of condition formulae of SDS^{OG} , $\phi \in \mathcal{F}(C, \mathcal{C})$, and $\mathcal{M}(C) = (\mathcal{M}_{R_1}, \mathcal{M}_{R_2}, \dots, \mathcal{M}_{R_l})$ be a condition attribute value meaning map associated with C from SDS^{OG} . The meaning $\|\phi\|_{SDS}^{\mathcal{M}(C)}$ of ϕ in SDS^{OG} with respect to $\mathcal{M}(C)$ is the set of all objects in U with the property ϕ defined inductively as follows:

- $\|(a, v)\|_{SDS}^{\mathcal{M}(C)} = \{u \in U : \models \mathcal{M}_{R_a}(a(u), v)\}$, where \mathcal{M}_{R_a} is a meaning of condition attribute values assigned to the attribute a in $\mathcal{M}(C)$,
- $\|\phi \wedge \phi'\|_{SDS}^{\mathcal{M}(C)} = \|\phi\|_{SDS}^{\mathcal{M}(C)} \cap \|\phi'\|_{SDS}^{\mathcal{M}(C)}$,
- $\|\phi \vee \phi'\|_{SDS}^{\mathcal{M}(C)} = \|\phi\|_{SDS}^{\mathcal{M}(C)} \cup \|\phi'\|_{SDS}^{\mathcal{M}(C)}$,
- $\|\neg \phi\|_{SDS}^{\mathcal{M}(C)} = U - \|\phi\|_{SDS}^{\mathcal{M}(C)}$.

We define the meaning of decision formulae in simple decision systems over ontological graphs in the standard way, i.e., as in the classic decision systems (cf. [11]).

Definition 9. Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system over ontological graphs, $\mathcal{F}(D, V_d)$ be the set of decision formulae of SDS^{OG} , $\psi \in \mathcal{F}(D, V_d)$. The meaning $\|\psi\|_{SDS}$ of ψ in SDS^{OG} is the set of all objects in U with the property ψ defined inductively as follows:

- $\|(a, v)\|_{SDS} = \{u \in U : a(u) = v\}$,
- $\|\phi \wedge \phi'\|_{SDS} = \|\phi\|_{SDS} \cap \|\phi'\|_{SDS}$,

- $\|\phi \vee \phi'\|_{SDS} = \|\phi\|_{SDS} \cup \|\phi'\|_{SDS}$,
- $\|\neg\phi\|_{SDS} = U - \|\phi\|_{SDS}$.

Definition 10. Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system over ontological graphs, $\mathcal{F}(C, \mathcal{C})$ and $\mathcal{F}(D, V_d)$ be the sets of condition formulae and decision formulae of SDS^{OG} , respectively, and $\mathcal{M}(C) = (\mathcal{M}_{R_1}, \mathcal{M}_{R_2}, \dots, \mathcal{M}_{R_l})$ be a condition attribute value meaning map associated with C from SDS^{OG} . A decision rule in SDS^{OG} with respect to $\mathcal{M}(C)$ is any expression of the form $\phi \Rightarrow \psi$, where $\phi \in \mathcal{F}(C, \mathcal{C})$, $\psi \in \mathcal{F}(D, V_d)$, and $\|\phi\|_{SDS}^{\mathcal{M}(C)} \neq \emptyset$.

A decision rule is true (valid, certain) in SDS^{OG} with respect to the condition attribute value meaning map $\mathcal{M}(C)$ associated with C from SDS^{OG} if and only if $\|\phi\|_{SDS}^{\mathcal{M}(C)} \subseteq \|\psi\|_{SDS}$.

Example 4. Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system given in Example 1. In this example, we will examine exemplary rules in SDS^{OG} with respect to different condition attribute value meaning maps associated with C from SDS^{OG} . Our task will be to determine whether a given decision rule is true (valid) in SDS^{OG} with respect to a selected condition attribute value meaning map.

1. Let us consider the condition attribute value meaning map in the form $\mathcal{M}(C) = (\mathcal{M}_=, \mathcal{M}_\leq, \mathcal{M}_\geq)$, i.e., all condition attributes are considered in their exact meanings. The following rule is true (valid) in SDS^{OG} :

$$(Dwelling, Cottage) \wedge (Job, Research worker) \Rightarrow \\ (Material Status, Medium),$$

because:

- $\|(Dwelling, Cottage) \wedge (Job, Research worker)\|_{SDS}^{\mathcal{M}(C)} = \{u_8\}$,
- $\|(Material Status, Medium)\|_{SDS}^{\mathcal{M}(C)} = \{u_1, u_3, u_7, u_8\}$.

2. Let us consider attribute values for *Dwelling* at the more general (abstract) level and let us take into consideration also synonyms for *Job*, i.e., the condition attribute value meaning map has the form $\mathcal{M}(C) = (\mathcal{M}_=, \mathcal{M}_\triangleleft^1, \mathcal{M}_\sim)$. The following rule in SDS^{OG} :

$$(Dwelling, Cottage) \wedge (Job, Research worker) \Rightarrow \\ (Material Status, Medium),$$

is not true (not valid) because:

- $\|(Dwelling, Cottage) \wedge (Job, Research worker)\|_{SDS}^{\mathcal{M}(C)} = \{u_5, u_8\}$ in view of
 - $(Cottage, Bungalow) \in GMR^1(Dwelling)$ (*Cottage* and *Bungalow* are generalized by *House* at the first level in the ontological graph $OG_{Dwelling}$ associated with the attribute *Dwelling*),
 - $(Research worker, Academic) \in SMR(Job)$ (*Research worker* and *Academic* are synonyms with respect to the ontological graph OG_{Job} associated with the attribute *Job*),

- $\|(\text{Material Status}, \text{Medium})\|_{SDS}^{\mathcal{M}(C)} = \{u_1, u_3, u_7, u_8\}$.

3. Let us consider the following rule in SDS^{OG} :

$$(\text{Vehicle}, \text{Car}) \wedge (\text{Dwelling}, \text{Flat}) \Rightarrow (\text{Material Status}, \text{Medium}).$$

Let us answer a question whether this rule is true (valid) in SDS^{OG} . The answer is not unique. In the presented approach, it depends on the assumed condition attribute value meaning map $\mathcal{M}(C)$. Assuming $\mathcal{M}(C) = (\mathcal{M}_=, \mathcal{M}_=, \mathcal{M}_=)$, a rule is true (valid) because:

- $\|(\text{Vehicle}, \text{Car}) \wedge (\text{Dwelling}, \text{Flat})\|_{SDS}^{\mathcal{M}(C)} = \{u_3\}$,
- $\|(\text{Material Status}, \text{Medium})\|_{SDS}^{\mathcal{M}(C)} = \{u_1, u_3, u_7, u_8\}$.

A rule becomes not true (not valid) if we assume more general meaning of attribute values, i.e., $\mathcal{M}(C) = (\mathcal{M}_{\triangleleft}, \mathcal{M}_{\triangleleft}, \mathcal{M}_{\triangleleft})$, because:

- $\|(\text{Vehicle}, \text{Car}) \wedge (\text{Dwelling}, \text{Flat})\|_{SDS}^{\mathcal{M}(C)} = \{u_1, u_3, u_6, u_7\}$ in view of:
 - $(\text{Lodging}, \text{House}) \in GMR^1(\text{Dwelling})$,
 - $(\text{Condominium}, \text{House}) \in GMR^1(\text{Dwelling})$,
- $\|(\text{Material Status}, \text{Medium})\|_{SDS}^{\mathcal{M}(C)} = \{u_1, u_3, u_7, u_8\}$.

Example 4 shows that, in simple decision systems over ontological graphs, the validity of a rule depends on the viewpoint from which attribute values (their meanings) are seen. For example, synonymy or generality of meanings of attribute values change a view on a decision system, and, among others, a view on the validity of rules in decision systems. This is a very important problem in data mining and data analysis. If we consider literally attribute values, then our solutions are sensitive to different meaning mistakes, uncertainties, vagueness, or ambiguities of data.

5 Conclusions and Further Works

The approach presented in this paper generalizes a look at decision rules in decision systems if some new knowledge describing semantic meanings of attribute values is available. We can distinguish several directions for further work. The first problem is validity of rules in simple decision systems over ontological graphs with respect to more sophisticated semantic relations between attribute values (concepts). There are a lot of different semantic relations between concepts defined in the literature (cf. [13]). Another task is to elaborate rule generation algorithms in simple decision systems over ontological graphs. The first thing to do is to consider the algorithm for generation of minimal decision rules true in simple decision systems over ontological graphs based on well known discernibility matrices in rough set theory (cf. [12]).

References

1. Gomez-Perez, A., Fernandez-Lopez, M., Corcho, O.: *Ontological Engineering*. Springer, London (2004)
2. Greco, S., Matarazzo, B., Słowiński, R.: Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research* 129(1), 1–47 (2001)
3. Ishizu, S., Gehrmann, A., Nagai, Y., Inukai, Y.: Rough ontology: Extension of ontologies by rough sets. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4557, pp. 456–462. Springer, Heidelberg (2007)
4. Köhler, J., Philipp, S., Specht, M., Rüegg, A.: Ontology based text indexing and querying for the semantic web. *Knowledge-Based Systems* 19, 744–754 (2006)
5. Midelfart, H., Komorowski, J.: A rough set framework for learning in a directed acyclic graph. In: Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N. (eds.) RSCTC 2002. LNCS (LNAI), vol. 2475, pp. 144–155. Springer, Heidelberg (2002)
6. Milstead, J.L.: Standards for relationships between subject indexing terms. In: Bean, C.A., Green, R. (eds.) *Relationships in the Organization of Knowledge*, pp. 53–66. Kluwer Academic Publishers, Dordrecht (2001)
7. Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., Swartout, W.: Enabling technology for knowledge sharing. *AI Magazine* 12(3), 36–56 (1991)
8. Pancerz, K.: Dominance-based rough set approach for decision systems over ontological graphs. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) *Proceedings of the FedCSIS 2012*, Wrocław, Poland, pp. 323–330 (2012)
9. Pancerz, K.: Toward information systems over ontological graphs. In: Yao, J., Yang, Y., Słowiński, R., Greco, S., Li, H., Mitra, S., Polkowski, L. (eds.) RSCTC 2012. LNCS (LNAI), vol. 7413, pp. 243–248. Springer, Heidelberg (2012)
10. Pancerz, K., Grochowalski, P.: Matching ontological subgraphs to concepts: a preliminary rough set approach. In: Hassanien, A., et al. (eds.) *Proceedings of the ISDA 2010*, Cairo, Egypt, pp. 1394–1399 (2010)
11. Pawlak, Z.: *Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht (1991)
12. Skowron, A., Rauszer, C.M.: The discernibility matrices and functions in information systems. In: Słowiński, R.W. (ed.) *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*, pp. 331–362. Kluwer Academic Publishers, Dordrecht (1992)
13. Storey, V.C.: Understanding semantic relationships. *The VLDB Journal* 2, 455–488 (1993)

Nonlinear Extension of the IRLS Classifier Using Clustering with Pairs of Prototypes

Michał Jezewski and Jacek M. Leski

Abstract. Classification is the primal problem in pattern recognition. The quadratic loss function is not good approximation of the misclassification error. In the presented paper a nonlinear extension of the IRLS classifier, which uses different loss functions, is proposed. The extension is done by means of fuzzy if-then rules. The fuzzy clustering with pairs of prototypes is applied to establish rules parameters values. Classification quality and computing time obtained for six benchmark databases is compared with the Lagrangian SVM method.

Keywords: Fuzzy Clustering, Rule-Base Classifiers, IRLS.

1 Introduction

Classification is the primal problem in pattern recognition. There are many classifiers: statistical, linear discriminant, k -nearest neighbor, kernel, neural network, classification tree and many more [1], [7], [8], [10]. The Support Vector Machine (SVM) method [9] is regarded as a one of the best. However, it is characterized by high computational cost. Therefore, an alternative methods like Lagrangian SVM [6] were proposed. Real data are characterized by noise and outliers. As a result classifier should be robust. Quadratic loss function is not good approximation of the

Michał Jezewski

Institute of Electronics, Silesian University of Technology, Akademicka 16,
44-100 Gliwice, Poland

e-mail: m.jezewski@polsl.pl

Jacek M. Leski

Institute of Medical Technology and Equipment, Roosevelt St. 118A, 41-800 Zabrze, Poland
Institute of Electronics, Silesian University of Technology, Akademicka 16,
44-100 Gliwice, Poland

e-mail: jacekl@itam.zabrze.pl, jleski@polsl.pl

misclassification error - it does not lead to a robust method. A new linear classifier design method based on regularized Iteratively Reweighted Least Squares (IRLS) criterion function was proposed [4]. It uses different loss functions. In [5], fuzzy clustering algorithm dedicated to classification methods is described. The goal of the presented paper is to apply it to nonlinear extension of the IRLS classifier.

2 Classifier Design Method

2.1 Fuzzy Clustering with Pairs of Prototypes

A fuzzy clustering algorithm dedicated to classification methods was proposed [5]. Its goal is to find pairs of prototypes from both classes of objects. In a given class prototypes should be located close to boundary with the second class. Final prototypes, which should be located near boundary between classes, are determined basing on obtained pairs. The algorithm is based on minimization of the following criterion function

$$J(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{V}^{(1)}, \mathbf{V}^{(2)}) = \sum_{i=1}^c \sum_{k \in \Omega_1} \left(u_{ik}^{(1)} \right)^m d_{ik}^2 + \sum_{i=1}^c \sum_{k \in \Omega_2} \left(u_{ik}^{(2)} \right)^m d_{ik}^2 + \eta \sum_{i=1}^c \left\| \mathbf{v}_i^{(1)} - \mathbf{v}_i^{(2)} \right\|, \quad (1)$$

with the constraints

$$\forall_{k \in \Omega_1} \sum_{i=1}^c u_{ik}^{(1)} = 1, \quad \forall_{k \in \Omega_2} \sum_{i=1}^c u_{ik}^{(2)} = 1, \quad (2)$$

where $\Omega_1 = \{k | \mathbf{x}_k \in \omega_1\}$ and $\Omega_2 = \{k | \mathbf{x}_k \in \omega_2\}$. The \mathbf{U} (\mathbf{V}) denote partition (prototype) matrix, c denotes number of clusters, \mathbf{x}_k represents k th object, m influences a fuzziness of clusters (usually $m = 2$ is chosen and such value was assumed), d_{ik} denotes the Euclidean distance between the i th prototype and the k th object. Two first components of the criterion represent classic fuzzy c -means algorithm and are responsible for clustering in both classes - upper indexes (1) and (2) denote the first (ω_1) and second (ω_2) class. The third component ensures minimization of Euclidean distances between prototypes in pairs - prototypes making pairs should move closer to each other and should be located near boundaries. The parameter η determines the proportion between clustering and minimizing distances between prototypes. The necessary conditions for minimization of the criterion are as follows:

$$\forall_{1 \leq s \leq c} \forall_{k \in \Omega_1} u_{sk}^{(1)} = \frac{(d_{sk})^{\frac{2}{1-m}}}{\sum_{j=1}^c (d_{jk})^{\frac{2}{1-m}}}; \quad \forall_{k \in \Omega_2} u_{sk}^{(2)} = \frac{(d_{sk})^{\frac{2}{1-m}}}{\sum_{j=1}^c (d_{jk})^{\frac{2}{1-m}}}, \quad (3)$$

$$\forall_{1 \leq s \leq c} \quad \mathbf{v}_s^{(1)} = \frac{\eta \sum_{k \in \Omega_1 \cup \Omega_2} (u_{sk})^m \mathbf{x}_k + \sum_{k \in \Omega_2} (u_{sk}^{(2)})^m \sum_{k \in \Omega_1} (u_{sk}^{(1)})^m \mathbf{x}_k}{\eta \sum_{k \in \Omega_1 \cup \Omega_2} (u_{sk})^m + \sum_{k \in \Omega_2} (u_{sk}^{(2)})^m \sum_{k \in \Omega_1} (u_{sk}^{(1)})^m}, \quad (4)$$

$$\forall_{1 \leq s \leq c} \quad \mathbf{v}_s^{(2)} = \frac{\eta \sum_{k \in \Omega_1 \cup \Omega_2} (u_{sk})^m \mathbf{x}_k + \sum_{k \in \Omega_1} (u_{sk}^{(1)})^m \sum_{k \in \Omega_2} (u_{sk}^{(2)})^m \mathbf{x}_k}{\eta \sum_{k \in \Omega_1 \cup \Omega_2} (u_{sk})^m + \sum_{k \in \Omega_1} (u_{sk}^{(1)})^m \sum_{k \in \Omega_2} (u_{sk}^{(2)})^m}. \quad (5)$$

The criterion minimization procedure may be found in [5]. Final prototypes are determined basing on the following formula

$$\forall_{1 \leq s \leq c} \quad \mathbf{v}_s = \frac{\sum_{k \in \Omega_1} (u_{sk}^{(1)})^m \mathbf{x}_k + \sum_{k \in \Omega_2} (u_{sk}^{(2)})^m \mathbf{x}_k}{\sum_{k \in \Omega_1} (u_{sk}^{(1)})^m + \sum_{k \in \Omega_2} (u_{sk}^{(2)})^m}. \quad (6)$$

In both classes the clustering was started from prototypes evenly located on the diagonal determined by 95% of minimum and 95% of maximum values of features. In experiments iterations (maximum number equals to 500) were stopped as soon as the absolute value of the difference between successive values of the criterion was less than 10^{-6} .

2.2 Linear Classifier

The criterion function of the linear IRLS classifier for k th iteration has the following form [4]

$$\min_{\mathbf{w}^{(k)} \in \mathbb{R}^{t+1}} J^{(k)}(\mathbf{w}^{(k)}) \triangleq \frac{1}{2} (\mathbf{X}\mathbf{w}^{(k)} - \mathbf{1})^\top \mathbf{H}^{(k)} (\mathbf{X}\mathbf{w}^{(k)} - \mathbf{1}) + \frac{\tau}{2} (\tilde{\mathbf{w}}^{(k)})^\top \tilde{\mathbf{w}}^{(k)}, \quad (7)$$

where $\mathbf{w} = [\tilde{\mathbf{w}}^\top, w_0]^\top \in \mathbb{R}^{t+1}$ is the weight vector of linear discriminant function $d(\mathbf{x}_i) \triangleq \mathbf{w}^\top \mathbf{x}'_i = \tilde{\mathbf{w}}^\top \mathbf{x}_i + w_0$, $\mathbf{x}_i \in \mathbb{R}^t$ denotes i th object from the N element training subset, $\mathbf{x}'_i = [\mathbf{x}_i^\top, 1]^\top$, $\mathbf{1}$ is the vector with all entries equal to 1. The \mathbf{X} matrix is defined as follows

$$\mathbf{X}^\top \triangleq [\varphi_1 \mathbf{x}'_1, \varphi_2 \mathbf{x}'_2, \dots, \varphi_N \mathbf{x}'_N], \quad (8)$$

where φ_i (equals to +1 or -1) is the class label indicating an assignment of i th object to one of two classes. The matrix with weights $\mathbf{H}^{(k)} = \text{diag}(h_1^{(k)}, h_2^{(k)}, \dots, h_N^{(k)})$ is used for both asymmetrization (relaxation) and changing the loss function (approximation of the misclassification error). If we have

$$h_i^{(k)} = \begin{cases} 0, & e_i^{(k-1)} \geq 0, \\ 1, & e_i^{(k-1)} < 0, \end{cases} \quad (9)$$

where

$$\mathbf{e}^{(k-1)} = \mathbf{X}\mathbf{w}^{(k-1)} - \mathbf{1}, \quad (10)$$

then obtained loss function may be called Asymmetric SQuaRe (ASQR). Changing the $h_i^{(k)}$ different loss functions may be obtained. Beside the ASQR, Asymmetric LINEar (ALIN), Asymmetric HUBer (AHUB), SIGmoidal (SIG), Asymmetric SIGmoidal-Linear (ASIGL), Asymmetric LOGarithmic (ALOG) and Asymmetric LOG-Linear (ALOGL) were proposed in [4]. All of them were applied in the presented work. In case of SIG and ASIGL $\alpha = 2.0$ was assumed. The second component of the criterion (7) is related to the statistical learning theory and ensures maximization of the margin of separation between both classes, parameter τ controls the proportion between both components.

2.3 Nonlinear Extension

An extension of the linear IRLS classifier to nonlinear classifier may be done by designing c linear classifiers, which are represented by c fuzzy if-then rules in the Takagi-Sugeno-Kang form [2], [3]. The rule base of the Takagi-Sugeno-Kang fuzzy system with linear functions in consequences consists of the following fuzzy if-then rules

$$\mathcal{R} = \left\{ \text{if } \bigwedge_{n=1}^t x_{kn} \text{ is } A_n^{(i)}, \text{ then } y = y^{(i)}(\mathbf{x}_k) = \mathbf{w}^{(i)\top} \mathbf{x}'_k \right\}_{i=1}^c, \quad (11)$$

where x_{kn} denotes input singleton, $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kt}]^\top$, c is number of rules, $A_n^{(i)}$ is the linguistic value of antecedent for the n th feature and the i th rule. The overall output for \mathbf{x}_k datum is obtained by the weighted average

$$y_0(\mathbf{x}_k) = \frac{\sum_{i=1}^c F^{(i)}(\mathbf{x}_k) y^{(i)}(\mathbf{x}_k)}{\sum_{j=1}^c F^{(j)}(\mathbf{x}_k)}, \quad (12)$$

where $F^{(i)}(\mathbf{x}_k)$ is the firing strength of the i th rule for the \mathbf{x}_k object. Gaussian memberships functions in the antecedents of if-then rules and the algebraic product as t -norm were applied. Then the firing strength may be written in the form

$$F^{(i)}(\mathbf{x}_k) = \exp \left[-\frac{1}{2} \sum_{n=1}^t \frac{(x_{kn} - v_{in})^2}{(s_{in})^2} \right]. \quad (13)$$

where s_{in} is a dispersion for the n th feature in antecedent of the i th rule. The final prototypes (6) were used as centers of Gaussian functions. The dispersions are calculated with the following formula

$$(s_{in})^2 = \frac{\sum_{k \in \Omega_1} \left(u_{ik}^{(1)} \right)^m (x_{kn} - v_{in})^2 + \sum_{k \in \Omega_2} \left(u_{ik}^{(2)} \right)^m (x_{kn} - v_{in})^2}{\sum_{k \in \Omega_1} \left(u_{ik}^{(1)} \right)^m + \sum_{k \in \Omega_2} \left(u_{ik}^{(2)} \right)^m}. \quad (14)$$

The normalized firing strength is determined by the formula

$$\overline{F^{(i)}}(\mathbf{x}_k) = \frac{F^{(i)}(\mathbf{x}_k)}{\sum_{j=1}^c F^{(j)}(\mathbf{x}_k)}. \quad (15)$$

Defining \mathbf{X}_G matrix

$$\mathbf{X}_G = \begin{bmatrix} \varphi_1 \overline{F^{(1)}}(\mathbf{x}_1) \mathbf{x}_1'^\top & \cdots & \varphi_1 \overline{F^{(c)}}(\mathbf{x}_1) \mathbf{x}_1'^\top \\ \varphi_2 \overline{F^{(1)}}(\mathbf{x}_2) \mathbf{x}_2'^\top & \cdots & \varphi_2 \overline{F^{(c)}}(\mathbf{x}_2) \mathbf{x}_2'^\top \\ \vdots & \ddots & \vdots \\ \varphi_N \overline{F^{(1)}}(\mathbf{x}_N) \mathbf{x}_N'^\top & \cdots & \varphi_N \overline{F^{(c)}}(\mathbf{x}_N) \mathbf{x}_N'^\top \end{bmatrix}, \quad (16)$$

and vectors $\widetilde{\mathbf{W}} = [\widetilde{\mathbf{w}}^{(1)\top}, \widetilde{\mathbf{w}}^{(2)\top}, \dots, \widetilde{\mathbf{w}}^{(c)\top}]^\top$, $\mathbf{W} = [\mathbf{w}^{(1)\top}, \mathbf{w}^{(2)\top}, \dots, \mathbf{w}^{(c)\top}]^\top$, criterion equivalent to (7) may be defined:

$$\min_{\mathbf{W}^{(k)} \in \mathbb{R}^{c(t+1)}} J^{(k)}(\mathbf{W}^{(k)}) \triangleq \frac{1}{2} (\mathbf{X}_G \mathbf{W}^{(k)} - \mathbf{1})^\top \mathbf{H}^{(k)} (\mathbf{X}_G \mathbf{W}^{(k)} - \mathbf{1}) + \frac{\tau}{2} \left(\widetilde{\mathbf{W}}^{(k)} \right)^\top \widetilde{\mathbf{W}}^{(k)}. \quad (17)$$

The criterion minimization procedure using the conjugate gradient algorithm is presented in [4] and was applied in the work. The procedure consists of two nested loops, 500 was assumed as a maximum number of iterations for each of them. Inner (outer) iterations were stopped as soon as the Euclidean norm in a successive pair of \mathbf{W} vectors was less than 10^{-6} (10^{-2}).

3 Results

Classification quality was examined using benchmark databases: 'Banana' (Ban), 'Breast Cancer' (BreC), 'Diabetis' (Diab), 'Heart' (Hea), 'Thyroid' (Thy). They were obtained from <http://ida.first.fraunhofer.de/projects/bench> (obtained 02.08.2008), where each of them was represented by 100 pairs of training and

testing subsets. Another database was Ripley synthetic two-class problem (**Syn**) - <http://www.stats.ox.ac.uk/pub/PRNN> (obtained 14.06.1997). In that case, the union of original training and testing subsets was randomly divided into 100 pairs of training (250 objects) and testing (1000 objects) subsets [4]. Results were compared with the Lagrangian SVM method obtained from <http://www.cs.wisc.edu/dmi/lsvm/> (obtained 26.06.2004).

Ten first pairs of subsets were used to determine values of the clustering (η) and classification (c, τ) parameters. In the first step, for each number of rules ($c = 2, 3, \dots, 8$ was assumed) values of η and τ ensuring the lowest mean misclassification error and its standard deviation was chosen. Both η and τ were chosen from the set $\{0.1, 1.0, 2.0, \dots, 10.0\}$. The number of rules (with corresponding values of η and τ) was chosen in the second step (also basing on the mean misclassification error and its standard deviation). For established values of c , η and τ , final result was calculated using all 100 pairs of subsets. Such procedure was performed for all applied loss functions.

In case of **Ban**, the best result was achieved using AHUB loss function and $c = 8$, $\eta = 0.1$, $\tau = 0.1$ (see Table 1). Figure 1 presents results of the clustering into eight clusters ($\eta = 0.1$) of 68th training subset of **Ban** (for 68th testing subset the lowest misclassification error was achieved). Black and white squares represent both classes of objects. Prototypes in both classes, connected in pairs by dotted lines, are represented by black and gray circles. Black triangles denote final prototypes. Presented clustering was not performed as it was expected because of the low value of the η . However, such value led to the highest classification quality. Figure 2 presents results of the clustering of the same dataset and also into eight clusters, but with higher values of the η , equals to 5.0. In that case pairs of prototypes changed their location, prototypes (especially these represented by black circles) moved closer to boundaries, distances between prototypes making pairs are smaller. In opposite to Figure 1, all final prototypes (except one) are located near boundary between classes, one of them in the top of figure is covered with pair of prototypes.

Table 1 presents results of the first step of the procedure determining parameters values for **Ban** and AHUB loss function. With the increase of the number of rules, the decrease of misclassification error is observed. For each number of rules, $\tau = 0.1$ led to best result. The lowest value is for eight rules, therefore $c = 8$, $\eta = 0.1$, $\tau = 0.1$ were chosen to calculate final result.

Table 2 summarizes final results obtained for different loss functions in comparison with the LSVM method. Each cell contains: mean misclassification error (top), standard deviation (middle), computing time normalized to the LSVM computing time. In case of the LSVM, the Gaussian kernel (K) was applied, $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\chi \|\mathbf{x} - \mathbf{x}_i\|^2)$; $\chi \in \mathbb{R}_+$. Values of the LSVM parameters were also determined using ten first pairs of subsets, but standard deviation of the mean misclassification error was not took into consideration. The LSVM v parameter was searched within the range from 0.01 to 1 with a step 0.01, χ was chosen from the set $\{0.01, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50\}$. All experiments were run on Intel Core2 Duo E8400 3GHz (only one core used) with 3.25GB RAM, running Windows XP (Service Pack 3) and MATLAB environment.

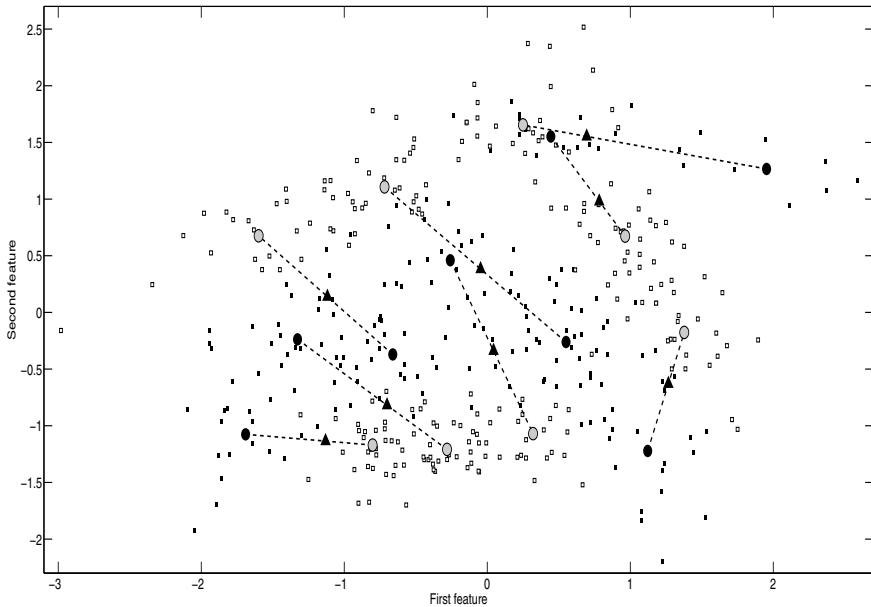


Fig. 1 An example of clustering results for Ban database; pairs of prototypes and final prototypes; $\eta = 0.1$

Table 1 Determination of parameters values, results of the first step (Ban, AHUB)

c	Misclassification error	η		τ
		Mean	St.Dev)	
2	28.443 (1.945)	2.0	0.1	
3	13.459 (2.525)	4.0	0.1	
4	12.426 (0.947)	0.1	0.1	
5	11.806 (0.693)	1.0	0.1	
6	11.312 (0.701)	0.1	0.1	
7	10.951 (0.898)	1.0	0.1	
8	10.829 (0.735)	0.1	0.1	

In case of three databases (Ban, BreC, Diab) the better result in relation to the LSVM was not achieved. For other three databases proposed classifier leads to lower mean misclassification error and its standard deviation - values which are the best and better in relation to the LSVM are written in bold in Table 2. Different loss functions lead to the lowest mean misclassification error - ASQR and AHUB for Ban, ALOGL for BreC and Thy, ALIN for Diab, ASIGL for Hea and Syn. In case of synthetic data (Ban, Syn), computing time was always (for all loss functions) lower with respect to the LSVM, especially in case of Ban (up to 33 times).

Table 2 Final results obtained for different loss functions and for the LSVM. Each cell contains: mean (top), standard deviation (middle) of misclassification error and normalized computing time (bottom). The best results for each database are in boldface.

	LSVM	ASQR	ALIN	SIG	ASIGL	AHUB	ALOG	ALOGL
Ban	10.340	10.880	10.972	11.338	11.041	10.880	10.883	10.888
	0.425	0.589	0.580	0.907	0.670	0.544	0.550	0.587
	1.00	0.03	0.09	0.05	0.06	0.06	0.06	0.08
BreC	25.195	26.805	26.831	26.844	26.896	27.013	26.857	26.740
	3.963	4.508	4.779	4.641	4.642	4.529	4.517	4.559
	1.00	0.89	4.56	1.16	1.59	1.51	1.61	3.15
Diab	23.143	23.423	23.213	23.807	23.307	23.267	23.303	23.353
	1.687	1.701	1.693	1.834	1.801	1.743	1.794	1.715
	1.00	0.72	3.28	0.59	1.31	1.10	1.16	2.20
Hea	15.680	16.060	15.670	16.000	15.550	15.960	15.610	16.380
	3.345	2.912	3.340	3.197	3.289	3.130	3.018	3.107
	1.00	1.93	5.99	1.58	3.22	2.44	2.58	3.08
Syn	9.460	9.566	9.620	9.576	9.423	9.490	9.643	9.550
	0.545	0.563	0.586	0.626	0.533	0.595	0.593	0.572
	1.00	0.27	0.54	0.14	0.26	0.35	0.38	0.86
Thy	4.147	3.427	4.933	4.173	4.493	3.813	3.267	3.227
	2.305	2.648	2.581	2.003	2.318	2.153	1.730	1.800
	1.00	14.60	3.46	1.27	1.95	12.11	10.36	42.32

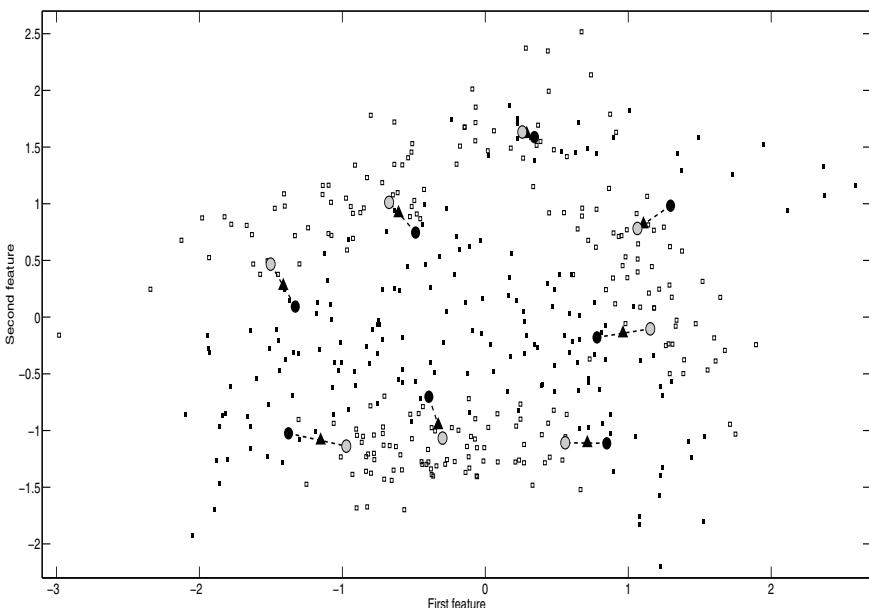


Fig. 2 An example of clustering results for **Ban** database; pairs of prototypes and final prototypes; $\eta = 5.0$

For the rest of databases, computing time was always (with the exception of three cases) higher, especially for Thy, up to 42 times. Taking into consideration different loss functions, ASQR has the lowest computing time in case of two databases, SIG in case of four. Among different loss functions, only in case of Ban and BreC number of rules ensuring the best results are similar (8 for Ban, 2 for BreC). Also only in case of these databases similar values of the η were chosen (0.1 for Ban, 1.0 for BreC). In case of the τ , only for synthetic data (Ban, Syn) any relation may be noticed - $\tau = 0.1$ was chosen.

Figure 3 presents 68th training subset of Ban and obtained discrimination curve and linear functions of eight if-then rules consequences (gray lines). Function of one consequence was located farther than Ban objects, so only seven lines are in the figure.

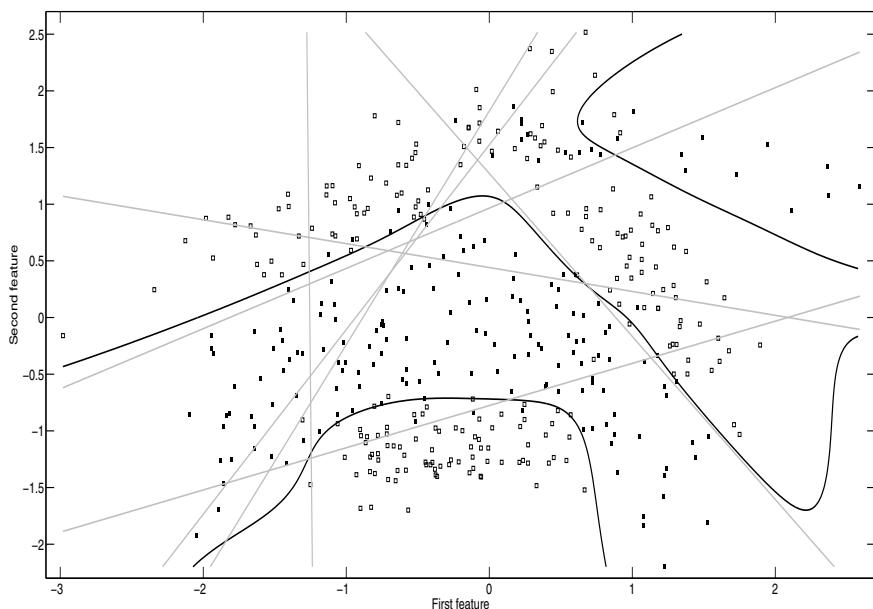


Fig. 3 Classification results for Ban database; discrimination curve and linear functions from consequences of rules

4 Conclusions

In the presented work it was shown, that the fuzzy clustering may be applied to establish fuzzy if-then rules parameters values to obtain nonlinear extension of the IRLS classifier. Seven different loss functions (approximations of the misclassification error) were used. Classification quality and computing time was examined using six benchmark databases and compared with the Lagrangian SVM method - modified version of the SVM, which is regarded as a one of the best classifiers. For

three databases lower mean misclassification error and its standard deviation were achieved. In case of two databases, for all loss functions computational time was lower. It is not possible to indicate the best loss function - different loss functions lead to the lowest mean misclassification error, depending on database.

References

1. Duda, R., Hart, P., Stork, D.: Pattern classification. John Wiley&Sons, New York (2001)
2. Leski, J.: An ε -margin nonlinear classifier based on if-then rules. IEEE Trans. Systems, Man and Cybernetics - Part B: Cybernetics 34(1), 68–76 (2004)
3. Leski, J.: Neuro-fuzzy systems. WNT, Warszawa (2008) (in Polish)
4. Leski, J.: Iteratively reweighted least squares classifier and its ℓ_2 - and ℓ_1 -regularized kernel versions. Bull. Pol. Ac.: Tech. 58(1), 171–182 (2010)
5. Leski, J., Jezewski, M.: Clustering algorithms for classification methods. Journal of Medical Informatics and Technologies 20, 11–18 (2012)
6. Mangasarian, O., Musicant, D.: Lagrangian support vector machines. Jou. Mach. Learn. Res. 1(1), 161–177 (2001)
7. Ripley, B.: Pattern recognition and neural networks. Cambridge University Press, Cambridge (1996)
8. Tou, J., Gonzalez, R.: Pattern recognition principles. Addison-Wesley, London (1974)
9. Vapnik, V.: Statistical learning theory. Wiley, New York (1998)
10. Webb, A.: Statistical pattern recognition. Arnold, London (1999)

Separable Linearization of Learning Sets by Ranked Layer of Radial Binary Classifiers

Leon Bobrowski and Magdalena Topczewska

Abstract. Layers of binary classifiers can be used in transformation of data sets composed of multivariate feature vectors. A new representation of data sets is obtained this way that depends on parameters of the classifiers in the layer. By a special, data driven choice of these parameters the ranked layer can be designed. The ranked layer has an important property of data sets linearization. It means that the data sets become linearly separable after transformation by ranked layer. The ranked layer can be built, *inter alia*, from radial or nearest neighbors binary classifiers.

1 Introduction

Classifiers are designed on the basis in a form of learning data sets according to variety of pattern recognition methods [1, 2]. Learning sets contain examples of objects assigned to selected categories (classes). Objects are typically represented in a standardized manner by a multivariate feature vectors. A binary classifier transforms multivariate feature vectors into a binary numbers equal to one or zero. A layer of binary classifiers transforms feature vectors into output vectors with binary components. A layer of binary classifiers aggregates input data sets if many feature vectors are transformed into the same output vector with binary components. The aggregation is separable if only feature vectors belonging to the same category may be merged into a single output vector. A layer aggregates leaning sets in linearly separable manner if each of the transformed sets can be separated from the others by a hyperplane in the transformed feature space. Linearly separable aggregation plays a special role in pattern recognition methods based on neural networks models [3]. In particular, the *Perceptron* model based on formal neurons is important in

Leon Bobrowski · Magdalena Topczewska
Faculty of Computer Science, Bialystok University of Technology,
ul. Wiejska 45A, Bialystok, Poland
e-mail: 1.bobrowski@pb.edu.pl

Leon Bobrowski
Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland

this context [4]. The linear separability of learning sets is an important issue also in the *support vector machines (SVM)* approach, one of the most popular methods in data mining, [5, 6]. An essential part of the *SVM* algorithms is the linear separability induction through kernel functions. The selection of the appropriate kernel functions is still an open and difficult problem in many practical applications of the support vector machines method. A family of K disjoined learning sets can be transformed into K linearly separable sets as a result of the transformation by ranked layer of formal neurons as it has been proved in the paper [7]. This result has been extended to the ranked layers of arbitrary binary classifiers in the work [8]. Designing ranked layers from radial binary classifiers is considered in the current work. In particularly, the computational procedure of ranked layer designing based on nearest neighbors is proposed.

2 Separable Learning Sets

Let us assume that each of m objects O_j ($j = 1, \dots, m$) is represented in a standardized manner by feature vectors $\mathbf{x}_j[n] = [x_{j1}, \dots, x_{jn}]^T$ belonging to a given n -dimensional feature space $F[n]$ ($\mathbf{x}_j[n] \in F[n]$). Each vector $\mathbf{x}_j[n]$ can be treated as a point of the n -dimensional feature space $F[n]$. Components x_{ji} of the feature vector $\mathbf{x}_j[n]$ can be a numerical results of n standardized examinations related to particular features x_i ($i = 1, \dots, n$) of a given object O_j ($x_{ji} \in \{0, 1\}$ or $x_{ji} \in R$). In practice, we can assume freely that the feature space $F[n]$ is equal to the n -dimensional real space R^n ($F[n] = R^n$). Let us assume that each object O_j may belong to one *category (class)* ω_k ($k = 1, \dots, K$). All the feature vectors $\mathbf{x}_j[n]$ that represent the objects O_j from one class ω_k can be collected as the k -th *learning set* C_k :

$$C_k = \{\mathbf{x}_j[n]: j \in J_k\} \quad (1)$$

where J_k is the set of indices j of objects O_j assigned to the k -th the class ω_k . The learning set C_k contains m_k feature vectors $\mathbf{x}_j[n]$ assigned to the k -th category ω_k . Assignment feature vectors $\mathbf{x}_j[n]$ to individual categories ω_k can be seen as a an additional knowledge about classification problem [1].

Definition 1. The learning sets C_k (1) are *separable* in the feature space $F[n]$, if they are disjoined in this space ($C_k \cap C_{k'} = \emptyset$, if $k \neq k'$). This means that the feature vectors $\mathbf{x}_j[n]$ and $\mathbf{x}_{j'}[n]$ belonging to different learning sets C_k and $C_{k'}$ cannot be equal:

$$(k \neq k') \Rightarrow (\forall j \in J_k) \text{ and } (\forall j \in J_{k'}) \mathbf{x}_j[n] \neq \mathbf{x}_{j'}[n] \quad (2)$$

We also take into consideration the separation of the learning sets C_k (1) sets by the hyperplanes $H(\mathbf{w}_k[n], \theta_k)$ in the feature space $F[n]$:

$$H(\mathbf{w}_k[n], \theta_k) = \{\mathbf{x}[n]: \mathbf{w}_k[n]^T \mathbf{x}[n] = \theta_k\}. \quad (3)$$

where $\mathbf{w}_k[n] = [w_{k1}, \dots, w_{kn}]^T \in R^n$ is the weight vector, $\theta_k \in R^1$ is the threshold, and $\mathbf{w}_k[n]^T \mathbf{x}[n]$ is the inner product.

Definition 2. The feature vector $\mathbf{x}_j[n]$ is situated on the *positive side* of the hyperplane $H(\mathbf{w}_k[n], \theta_k)$ (3) if and only if $\mathbf{w}_k[n]^T \mathbf{x}_j[n] > \theta_k$. Similarly, the vector $\mathbf{x}_j[n]$ is situated on the *negative side* of $H(\mathbf{w}_k[n], \theta_k)$ if and only if $\mathbf{w}_k[n]^T \mathbf{x}_j[n] < \theta_k$.

Definition 3. The learning sets (1) are *linearly separable* in the n -dimensional feature space $F[n]$ if each of the sets C_k can be fully separated from the sum of the remaining sets C_i by some hyperplane $H(\mathbf{w}_k[n], \theta_k)$ (3):

$$(\forall k \in \{1, \dots, K\}) (\exists \mathbf{w}_k[n], \theta_k) \quad (\forall \mathbf{x}_j[n] \in C_k) \quad \mathbf{w}_k[n]^T \mathbf{x}_j[n] > \theta_k \\ \text{and } (\forall \mathbf{x}_j[n] \in C_i, i \neq k) \quad \mathbf{w}_k[n]^T \mathbf{x}_j[n] < \theta_k \quad (4)$$

In accordance with the inequalities (4), all the vectors $\mathbf{x}_j[n]$ from the learning set C_k are situated on the positive side of the hyperplane $H(\mathbf{w}_k[n], \theta_k)$ (3) and all the vectors $\mathbf{x}_j[n]$ from the remaining sets C_i are situated on the negative side of this hyperplane.

3 Radial Binary Classifiers

The binary classifiers $BC_i(\mathbf{v}_i[N])$ operates on input (feature) vectors $\mathbf{x}[n]$ ($\mathbf{x}[n] \in F[n]$) and are characterized by such a decision rule $r_i(\mathbf{v}_i[N]; \mathbf{x}[n])$ which depends on the parameter vector $\mathbf{v}_i[N] = [v_{i1}, \dots, v_{iN}]^T$ in the parameter space R^N ($\mathbf{v}_i[N] \in R^N$) [1]. The vector $\mathbf{v}_i[N]$ is composed of N parameters \mathbf{v}_{ik} ($\mathbf{v}_{ik} \in R$). Each classifier $BC_i(\mathbf{v}_i[N])$ has the binary output $r_i = r_i(\mathbf{v}_i[N]; \mathbf{x}[n])$ ($r_i \in \{0, 1\}$, $i = 1, \dots, L$).

Definition 4. The *activation field* $A_i(\mathbf{v}_i[N])$ of the i -th binary classifier $BC_i(\mathbf{v}_i[N])$ with the decision rule $r_i(\mathbf{v}_i[N]; \mathbf{x}[n])$ is the set of such input vectors $\mathbf{x}[n]$ which activate this classifier.

$$A_i(\mathbf{v}_i[N]) = \{\mathbf{x}[n] : r_i(\mathbf{v}_i[N]; \mathbf{x}[n]) = 1\} \quad (5)$$

The radial binary classifier $RC(\mathbf{w}_i[n], \rho_i)$ has the activation field $A_i(\mathbf{w}_i[n], \rho_i)$ (5) in a form of a sphere with the *center* $\mathbf{w}_i[n] = [w_{i1}, \dots, w_{in}]^T$ and the *radius* ρ_i ($\rho_i > 0$) [1]. The radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ can be characterized by the below decision rule $r_{RC}(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ which is based on the distances $\delta(\mathbf{w}_i[n], \mathbf{x}[n])$ between the point $\mathbf{x}[n]$ and the center $\mathbf{w}_i[n]$:

$$r_{RC}(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n]) = \begin{cases} 1 & \text{if } (\mathbf{w}_i[n], \mathbf{x}[n]) \leq \rho_i \\ 0 & \text{if } (\mathbf{w}_i[n], \mathbf{x}[n]) > \rho_i \end{cases} \quad (6)$$

In accordance with the above decision rule, the radial classifier $RC(\mathbf{w}_i[n], \rho_i)$ is activated by the input vector $\mathbf{x}[n]$ ($r_{RC}(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n]) = 1$) if and only if the distance $(\mathbf{w}_i[n], \mathbf{x}[n])$ between the vector $\mathbf{x}[n]$ and the center $\mathbf{w}_i[n]$ is no greater than the radius ρ_i . The decision rule $r_{RC}(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ (6) of the radial classifier $RC(\mathbf{w}_i[n], \rho_i)$

depends on the $n + 1$ parameters $\mathbf{w}_i[n]$ and ρ_i . The shape of the activation field $A_{RC}(\mathbf{w}_i[n], \rho_i)$ (5) depends of the choice of the distance function $\delta(\mathbf{w}_i[n], \mathbf{x}[n])$. A few examples of distance functions $\delta(\mathbf{w}_i[n], \mathbf{x}[n])$ are given below [9]:

$$\begin{aligned}\delta_E(\mathbf{w}_i[n], \mathbf{x}[n]) &= ((\mathbf{x}[n] - \mathbf{w}_i[n])^T (\mathbf{x}[n] - \mathbf{w}_i[n]))^{\frac{1}{2}} && \text{the Euclidean dist.} \\ \delta_{L_1}(\mathbf{w}_i[n], \mathbf{x}[n]) &= \sum_{l=1, \dots, N} |w_{il} - x_l| && \text{the } L_1 \text{ distance} \\ \delta_M(\mathbf{w}_i[n], \mathbf{x}[n]) &= ((\mathbf{x}[n] - \mathbf{w}_i[n])^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}[n] - \mathbf{w}_i[n]))^{\frac{1}{2}} && \text{the Mahalanobis dist.}\end{aligned}\quad (7)$$

where $\boldsymbol{\Sigma}$ is the covariance matrix $n \times n$ designed on the base of m feature vectors $x_j[n]$. The shapes of the activation fields $A_{RC}(\mathbf{w}_i[n], \rho_i)$ (5) with these distance functions $\delta(\mathbf{w}_i[n], \mathbf{x}[n])$ are sketch on the Fig. 1.

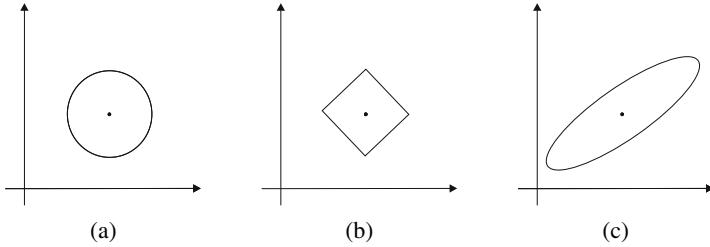


Fig. 1 The shapes of the activation fields $A_{RC}(\mathbf{w}_i[n], \rho_i)$ (5) with different distance functions $\delta(\mathbf{w}_i[n], \mathbf{x}[n])$ (7)

4 Separable Layers of Binary Classifiers

The layer composed of L radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ with the decision rules $r(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ (6) produces the output vectors $\mathbf{r}[L]$ with L binary components r_i ($r_i \in \{0, 1\}$):

$$\mathbf{r}[L] = [r_1, \dots, r_L]^T = [r(\mathbf{w}_1[n], \rho_1; \mathbf{x}[n]), \dots, r(\mathbf{w}_L[n], \rho_L; \mathbf{x}[n])]^T \quad (8)$$

The layer of L binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ transforms feature vectors $\mathbf{x}_j[n]$ from the learning sets C_k (1) into the sets R_k of the binary output vectors $\mathbf{r}_j[L]$:

$$R_k = \{\mathbf{r}_j[L]: j \in I_k\} \quad (9)$$

where

$$(\forall j \in \{1, \dots, m\}) \mathbf{r}_j[L] = [r(\mathbf{w}_1[n], \rho_1; \mathbf{x}_j[n]), \dots, r(\mathbf{w}_L[n], \rho_L; \mathbf{x}_j[n])]^T \quad (10)$$

Definition 5. The layer of L binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) is separable, if it preserves separability (2) of the learning sets C_k once they are transformed into the sets R_k (9). It means that the below implication is preserved after the transformation (10) by the layer:

$$(k \neq k') \Rightarrow (\forall j \in J_k) \text{ and } (j' \in J_{k'}) \mathbf{r}_j[L] \neq \mathbf{r}_{j'}[L] \quad (11)$$

Definition 6. The layer of L binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) is linearly separable, if the separable learning sets C_k (2) become linearly separable sets R_k (9) after the transformation (10) by this layer:

$$\begin{aligned} (\forall k \in \{1, \dots, K\}) \ (\exists \mathbf{w}_k[L], \theta_k) \ (\forall \mathbf{r}_j[L] \in R_k) \ \mathbf{w}_k[L]^T \mathbf{r}_j[L] > \theta_k \\ \text{and } (\forall \mathbf{r}_j[L] \in R_i, i \neq k) \ \mathbf{w}_k[L]^T \mathbf{r}_j[L] < \theta_k \end{aligned} \quad (12)$$

Let us remark that each linearly separable layer of binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) is also a separable layer. It has been proved that the ranked layer of binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) is linearly separable [8].

5 Designing Ranked Layers of Radial Binary Classifiers

Designing ranked layers can be based on a search for *homogeneous* active fields $A_i(\mathbf{w}_i[n], \rho_i)$ (5) of radial classifiers $RC(\mathbf{w}_i[n], \rho_i)$ with the binary decision rules $r_{RC}(\mathbf{w}_i[n], \rho_i; \mathbf{x}[n])$ (6):

$$A_i(\mathbf{w}_i[n], \rho_i) = \{\mathbf{x}_j[n]: r_{RC}(\mathbf{w}_i[n], \rho_i; \mathbf{x}_j[n]) = 1\} \quad (13)$$

Definition 7. The active field $A_i(\mathbf{w}_i[n], \rho_i)$ (13) is *homogeneous* in respect to the learning sets C_k (2) if it contains feature vectors $\mathbf{x}_j[n]$ belonging to only one of the sets C_k .

The active field $A_i(\mathbf{w}_i[n], \rho_i)$ (13) is not *inhomogeneous* if it contains feature vectors $\mathbf{x}_j[n]$ belonging to more than one learning sets C_k (2). In order to achieve a high *generalization power* of the layer the below designing postulate concerning the homogeneous active field $A_i(\mathbf{w}_i[n], \rho_i)$ (13) was introduced [8]:

Designing postulate I: The homogeneous active field $A_i(\mathbf{w}_i[n], \rho_i)$ (13)
should contain a lot of feature vectors $\mathbf{x}_j[n]$ belonging
to only one learning set C_k (2). (14)

In order to meet the postulate (14) the maximal homogeneous sphere S_j is computed for each of m feature vectors $\mathbf{x}_j[n]$. The sphere S_j is defined as the homogeneous active field $A_j(\mathbf{x}_j[n], \rho_j)$ (14) centered in the point $\mathbf{x}_j[n]$:

$$S_j = \max_{\rho} A_j(\mathbf{x}_j[n], \rho_j) \quad (15)$$

where the above symbol $\max_{\rho} A_j(\mathbf{x}_j[n], \rho_j)$ means such homogeneous active field $A_j(\mathbf{x}_j[n], \rho_j)$ (14) which has the largest radius ρ_j . The homogeneous sphere S_j (15) should be characterized by the maximal number M_j of feature vectors $\mathbf{x}_j[n]$ from one learning set C_k (2) ($\mathbf{x}_j[n] \in C_k$) contained in S_j . The *optimal sphere* S_{j*} (15) is selected on the basis of the numbers M_j of feature vectors $\mathbf{x}_j[n]$ from one learning set C_k (2) ($\mathbf{x}_j[n] \in C_k$) in particular areas S_j (15). The optimal sphere S_{j*} (15), is characterized by the highest number M_{j*} of feature vectors $\mathbf{x}_j[n]$:

$$(\forall j \in \{1, \dots, m\}) M_{j*} \geq M_j \quad (16)$$

Let us describe the multistage procedure of the ranked layer designing from radial classifiers $RC(\mathbf{w}_i[n], \rho_i)$ with the binary decision rules $r_{RC}(\mathbf{w}_i[n], \rho_i; \mathbf{x}_i[n])$ (6):

Ranked layer designing procedure (17)

Stage i. (Start)

- put $l=1$ and define the sets $C_k(l)$: $(\exists k \in \{1, \dots, K\}) C_k(l) = C_k$ (1)

Stage ii. (Optimal homogeneous sphere S_{j} (15))*

- find parameters j^* , k^* , $\mathbf{x}_{j*}[n]$ and ρ_j of the optimal homogeneous sphere S_{j*} (15) and the reduced set C_k (1)

Stage iii. (Reduction of the set C_k (1))

- delay such feature vectors $\mathbf{x}_j[n]$ which are contained within the optimal sphere S_{j*} (15)

$$\begin{aligned} C_{k*}(l+1) &= C_{k*}(l) - \{\mathbf{x}_j[n] : \mathbf{x}_j[n] \in S_{j*} \text{ (15)}\} \\ \text{and } (\forall k \in \{1, \dots, K\} \text{ where } k \neq k^*) C_k(l+1) &= C_k(l) \end{aligned} \quad (18)$$

Stage iv. (Stop criterion)

if all the sets $C_k(l+1)$ are empty *then stop*

else increase the index l by one ($l \rightarrow l + 1$) *and go to the Stage ii.*

It can be seen that each radial binary classifier $RC(\mathbf{w}_i[n], \rho_i)$ (6) added to the layer reduces (18) the set $C_{k*}(l)$ by at least one feature vector $\mathbf{x}_{j*}[n]$. Based on this property, it can be proved that if the learning sets C_k (1) are separable (2), then after finite number L of steps the procedure will be stopped. The following Lemma results [8]:

Lemma 1. *The number L of radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) in the ranked layer is no less than the number K of the learning sets C_k (1) and no greater than the number m of the feature vectors $\mathbf{x}_j[n]$ in these sets.*

$$K \leq L \leq m \quad (19)$$

The minimal number $L = K$ of radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) appears in the ranked layer when whole learning sets C_k (1) are reduced (13) during successive steps l . The maximal number $L = m$ of elements appears in the ranked layer when only single elements $\mathbf{x}_j[n]$ are reduced during successive steps l .

Theorem 1. *The sets R_k (9) obtained in result of transformation (10) of separable learning sets C_k (2) by the ranked layer (17) of L radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) are linearly separable (4).*

The proof of the above theorem can be given by repeating the argumentation from the analysis given in the works [7] and [8]. The ranked layer design procedure (17) involves the generation of a sequence of the optimal homogeneous spheres S_j (15). The procedure (17) is stopped if each feature vector $\mathbf{x}_j[n]$ is located in one sphere S_j (15). The postulate (14) can be treated as an element of the greedy strategy aimed at the ranked layer designing with a great power of generalization. A more general designing postulate can be formulated as:

Designing postulate II: The ranked layer should include the minimal number

$$L \quad (19) \text{ of radial binary classifiers } RC(\mathbf{w}_i[n], \rho_i) \quad (6) \quad (20)$$

We can remark also, that the assumptions of the procedure (17) may be less restrictive in other points. First of all, demand that all spheres S_j (15) should be homogeneous, can be relaxed in some limits. Also, not every feature vector $\mathbf{x}_j[n]$ must be placed in certain sphere S_j (15). A limited number of vectors $\mathbf{x}_j[n]$ may remain beyond the spheres S_j (15). After this kind of the procedure (17) relaxation, the linear separability (12) of the sets R_k (9) is no longer guaranteed. The sets R_k (9) may become almost linearly separable [10]. A greater generalization power of the designed layer of binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) may be achieved by taking into account the sets R_k (9) which are not necessary linearly separable (14) but only almost linearly separable [8].

6 Experimental Results

To demonstrate the individual steps of the ranked layers' designing two experiments are presented. The first is a toy example with small amount of objects and the second is the well-known and well-understood problem of flowers classification.

6.1 Toy Example

In the experiment data set containing collection of objects presented in the Figure 2(a) was chosen. The objects belong to two classes - the label for ω_1 class is 1 and the circle is a marker specifier. Analogically, the label for ω_2 class is 2 and the cross is a marker specifier. The objects are not linearly separable. Application of the ranked layers of binary classifiers allowed to separate objects from examined two classes completely in three steps. The results are presented in the Table 1.

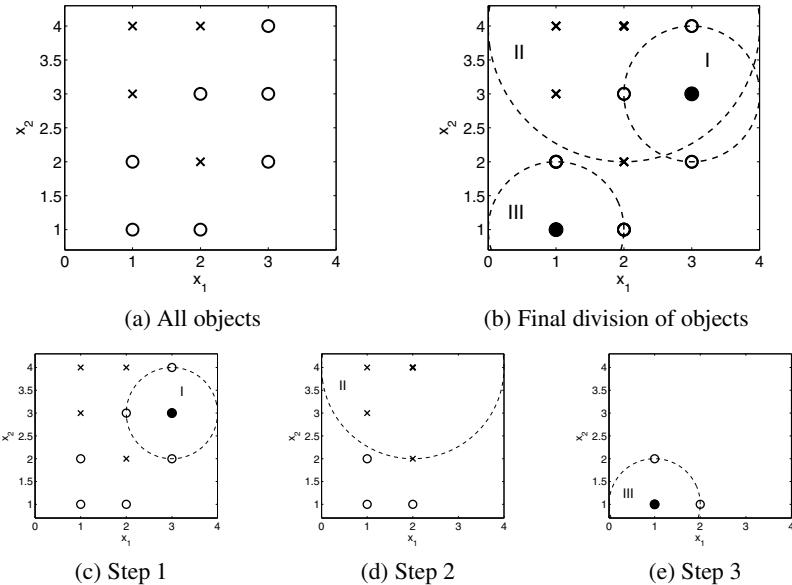


Fig. 2 Toy example data set: (a) a scatter plot; (b) final S_i circles; (c)-(e) steps of the classification procedure

Table 1 Results for the Toy example

index j	feature vector \mathbf{x}_j	transformed vector \mathbf{r}_j	category
1	[1,4]	[0,1,0]	cross
2	[2,4]	[0,1,0]	cross
3	[3,4]	[1,1,0]	circle
4	[1,3]	[0,1,0]	cross
5	[2,3]	[1,1,0]	circle
6	[3,3]	[1,1,0]	circle
7	[1,2]	[0,0,1]	circle
8	[2,2]	[0,1,0]	cross
9	[3,2]	[1,0,0]	circle
10	[1,1]	[0,0,1]	circle
11	[2,1]	[0,0,1]	circle

The S_1 sphere is centered at the [3,3] point and the radius is equaled $\rho_1 = 1.000$, the S_2 sphere is centered at the [2,4] point and the radius is $\rho_2 = 2.000$, and finally the S_3 sphere is centered at the [1,1] point and the radius is $\rho_3 = 1.000$.

The weight vectors of the hyperplanes separating the two transformed sets of objects have the forms: $\mathbf{v}_1 = [\frac{1}{2}, -\frac{1}{4}, \frac{1}{8}]^T$ and $\mathbf{v}_2 = [-\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}]^T$ respectively.

6.2 Experiment 2

In the second experiment the *Iris* data set was chosen. It is the well-known and well-understood problem of three species of irises, where each of 150 flowers is described by four attributes and belongs to one of three classes.

Table 2 Results for the Iris data set (\mathbf{x}_C - the center of the S_i sphere, ρ_i - the radius of the S_i sphere, m_i - number of classified objects by the S_i sphere)

step i	sphere center \mathbf{x}_{Ci}	radius ρ_i	m_i	category
1	[5.1,3.5,1.4,0.2]	2.0905	50	Iris-setosa
2	[5.7,2.6,3.5,1.0]	1.4629	39	Iris-versicolor
3	[7.2,3.6,6.1,2.5]	1.5684	35	Iris-virginica
4	[5.8,2.8,5.1,2.4]	0.7874	10	Iris-virginica
5	[6.6,2.9,4.6,1.3]	0.6245	9	Iris-versicolor
6	[5.9,3.2,4.8,1.8]	0.6481	2	Iris-versicolor
7	[4.9,2.5,4.5,1.7]	1.6583	5	Iris-virginica

The results are presented in the Table 2. In seven steps the classification of objects belonging to three classes was performed. In the first step the whole category *Iris-setosa* was perfectly classified by the S_1 sphere with the $\mathbf{x}_{C1} = [5.1, 3.5, 1.4, 0.2]$ center and the $\rho_1 = 2.0905$ radius. In the second step 39 objects belonging to the *Iris-versicolor* category were classified by the S_2 sphere with the $\mathbf{x}_{C2} = [5.7, 2.6, 3.5, 1.0]$ center and the $\rho_2 = 1.4626$ radius. In the next two steps, 35 and 10 objects belonging to the *Iris-virginica* category were classified by the S_3 and S_4 spheres ($\mathbf{x}_{C3} = [7.2, 3.6, 6.1, 2.5]$, $\rho_3 = 1.5684$ and $\mathbf{x}_{C4} = [5.8, 2.8, 5.1, 2.4]$, $\rho_4 = 0.7874$). The steps 5 and 6 concerned classification of 11 objects belonging to the *Iris-versicolor* category. Two spheres were built with the centers and radii: $\mathbf{x}_{C5} = [6.6, 2.9, 4.6, 1.3]$, $\rho_5 = 0.6245$ and $\mathbf{x}_{C6} = [5.9, 3.2, 4.8, 1.8]$, $\rho_6 = 0.6481$. Finally the S_7 sphere ($\mathbf{x}_{C7} = [6.6, 2.9, 4.6, 1.3]$, $\rho_7 = 0.6245$) classified 5 objects belonging to the *Iris-virginica* category.

The weight vectors of the hyperplanes $h_k = H(\mathbf{v}_k[7], 0) = \{\mathbf{r}[7] : \mathbf{v}_k[7]^T \mathbf{r}[7] = 0\}$ ($k = 1, 2, 3$) separating the transformed sets of objects have the forms: $\mathbf{v}_1[7] = [\frac{1}{2}, -\frac{1}{4}, -\frac{1}{8}, -\frac{1}{16}, -\frac{1}{32}, -\frac{1}{64}, -\frac{1}{128}]^T$, $\mathbf{v}_2[7] = [-\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}, -\frac{1}{16}, \frac{1}{32}, \frac{1}{64}, -\frac{1}{128}]^T$ and $\mathbf{v}_3[7] = [-\frac{1}{2}, -\frac{1}{4}, \frac{1}{8}, \frac{1}{16}, -\frac{1}{32}, -\frac{1}{64}, \frac{1}{128}]^T$.

7 Conclusions

The ranked layer of radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) allows to transform separable learning sets C_k (2) into linearly separable (4) sets R_k (9). The linearization of learning sets C_k (1) is particularly important in the context of the support vector

machines (SVM) [5]. The linearization (4) of the learning sets C_k (1) in the SVM approach is achieved, not always successfully, by introducing kernel functions.

The data driven procedure (17) of the ranked layer designing from radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) has been proposed in this paper. Implementation of the proposed procedure (17) can be relatively easy. Sophisticated computational techniques are not necessarily needed for an implementation of the basic form of the designing procedure (17). The basic procedure (17) can be modified in many ways.

The problem of designing ranked layers with a high generalization power still has a lot of unresolved issues. One of these issues is the optimization of the ranked layer built with different types of classifiers. For example, the designed layer may include both radial binary classifiers $RC(\mathbf{w}_i[n], \rho_i)$ (6) as well as formal neurons [8].

References

1. Duda, O.R., Hart, P.E., Stork, D.G.: *Pattern Classification*. J. Wiley, New York (1986)
2. Fukunaga, K.: *Introduction to Statistical Pattern Recognition*. Academic Press (1972)
3. Bishop, C.M.: *Pattern recognition and Machine Learning*. Springer (2006)
4. Rosenblatt, F.: *Principles of neurodynamics*. Spartan Books, Washington (1962)
5. Vapnik, V.N.: *Statistical Learning Theory*. J. Wiley, New York (1998)
6. Hand, D.J., Mannila, H., Smyth, P.: *Principles of Data Mining*. MIT Press (2001)
7. Bobrowski, L.: Design of piecewise linear classifiers from formal neurons by some basis exchange technique. *Pattern Recognition* 24(9), 863–870 (1991)
8. Bobrowski, L.: Induction of linear separability through the ranked layers of binary classifiers. In: Iliadis, L., Jayne, C. (eds.) *EANN/AIAI 2011, Part I. IFIP AICT*, vol. 363, pp. 69–77. Springer, Heidelberg (2011)
9. Johnson, R.A., Wichern, D.W.: *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., Englewood Cliffs (1991)
10. Bobrowski, L.: Almost separable data aggregation by layers of formal neurons. In: 13th International Conference: KDS 2007: Knowledge-Dialogue-Solution, Sofia, June 18–24 (2007); *International Journal on Information Theories and Applications*, pp. 34–41 (2007)

Hidden Markov Models For Two-Dimensional Data

Janusz Bobulski

Abstract. Hidden Markov models are well-known methods for image processing. They are used in many areas where 1D data are processed. In the case of 2D data, there appear some problems with application HMM. There are some solutions, but they convert input observation from 2D to 1D, or create parallel pseudo 2D HMM, which is set of 1D HMMs in fact. This paper describes authentic 2D HMM with two-dimensional input data, and its application for pattern recognition in image processing.

1 Introduction

Hidden Markov models (HMM) are widely apply in data classification. They are used in speech recognition, character recognition, biological sequence analysis, financial data processing, texture analysis, face recognition, etc. This widely application of HMM is result of its effectiveness. An extension of the HMM to work on two-dimensional data is 2D HMM. A 2D HMM can be regarded as a combination of one state matrix and one observation matrix, where transition between states take place according to a 2D Markovian probability and each observation is generated independently by the corresponding state at the same matrix position. It was noted that the complexity of estimating the parameters of a 2D HMMs or using them to perform maximum a posteriori classification is exponential in the size of data. Similar to 1D HMM, the most important thing for 2D HMMs is also to solve the three basic problems, namely, probability evolution, optimal state matrix and parameters estimation.

When we process one-dimensional data, we have good tools and solution for this. Unfortunately, this is unpractical in image processing, because the images are

Janusz Bobulski
Czestochowa University of Technology
Institute of Computer and Information Science
Dabrowskiego Street 73, 42-200 Czestochowa, Poland
e-mail: janusz@icis.pcza.pl

two-dimensional. When you convert an image from 2D to 1D, you lose some information. So, if we process two-dimensional data, we should apply two-dimensional HMM, and this 2D HMM should work with 2D data. One of solutions is pseudo 2D HMM [2, 6, 10]. This model is extension of classic 1D HMM. There are super-states, which mask one-dimensional hidden Markov models (Fig. 1). Linear model is the topology of superstates, where only self transition and transition to the following superstate are possible. Inside the superstates there are linear 1D HMM. The state sequences in the rows are independent of the state sequences of neighboring rows. Additional, input data are divided to the vector. So, we have 1D model with 1D data in practise.

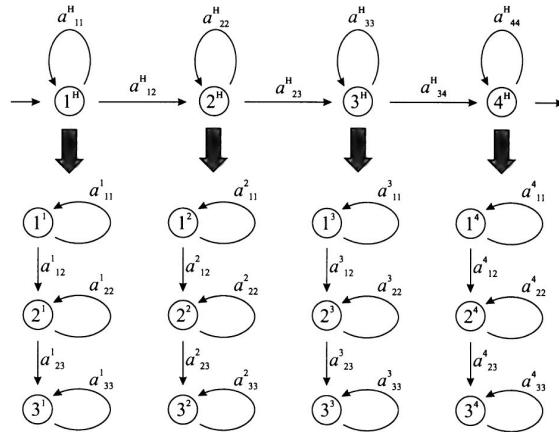


Fig. 1 Pseudo 2D HMM [1]

Other approach to image processing use two-dimensional data present in works [4] and [7]. The solutions base on Markov Random Fields (MRF) and give good results for classification and segmentation, but not in pattern recognition. Interesting results showed in paper [11]. This article presents analytic solution and evidence of correctness two-dimensional HMM. But this 2D HMM is similar to MRF, works with one-dimensional data and can be apply only for left-right type of HMM. This article presents real solution for 2D problem in HMM. There is show true 2D HMM which processes 2D data.

2 Classic 1D HMM

HMM is a double stochastic process with underlying stochastic process that is not observable (hidden), but can be observed through another set of stochastic processes that produce a sequence of observation [8]. Let $O = \{O_1, \dots, O_T\}$ be the sequence of

observation of feature vectors, where T is the total number of feature vectors in the sequence. The statistical parameters of the model may be defined as follows [5, 9]:

- The number of states of the model, N
- The number of symbols M
- The transition probabilities of the underlying Markov chain, $A = \{a_{ij}\}$, $1 \leq i, j \leq N$, where a_{ij} is the probability of transition from state i to state j
- The observation probabilities, $B = \{b_{jm}\}$ $1 \leq j \leq N, 1 \leq m \leq M$ which represents the probability of generate the m_{th} symbol in the j_{th} state.
- The initial probability vector, $\Pi = \{\pi_i\}$ $1 \leq i \leq N$.

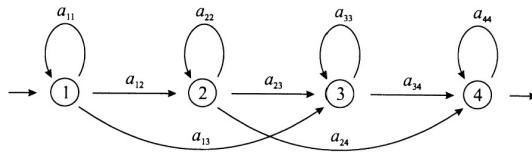


Fig. 2 One-dimensional HMM

Hence, the HMM requires three probability measures to be defined, A, B, Π and the notation $\lambda = (A, B, \Pi)$ is often used to indicate the set of parameters of the model. In the proposed method, one model is made for each part of the face. The parameters of the model are generated at random at the beginning. Then they are estimated with Baum-Welch algorithm, which is based on the forward-backward algorithm. The forward algorithm calculates the coefficient $\alpha_t(i)$ (probability of observing the partial sequence $(o_1, , o_t)$ such that state q_t is i). The backward algorithm calculates the coefficient $\beta_t(i)$ (probability of observing the partial sequence $(o_{t+1}, , o_T)$ such that state q_t is i). The Baum-Welch algorithm, which computes the λ , can be described as follows [5]:

1. Let initial model be λ_0
2. Compute new λ based on λ_0 and observation O
3. If $\log(P(O|\lambda) - \log(P(O)|\lambda_0) < \text{DELTA}$ stop
4. Else set $\lambda \rightarrow \lambda_0$ and go to step 2.

The parameters of new model λ , based on λ_0 and observation O , are estimated from equation of Baum-Welch algorithm [9], and then are recorded to the database.

3 Three Basic Problems

There are three fundamental problems of interest that must be solved for HMM to be useful in some applications. These problems are the following:

1. Given observation $O = (o_1, o_2, , o_T)$ and model $\lambda = (A, B, \Pi)$, efficiently compute $P(O|\lambda)$

2. Given observation $O = (o_1, o_2, \dots, o_T)$ and model λ find the optimal state sequence $q = (q_1, q_2, \dots, q_T)$
3. Given observation $O = (o_1, o_2, \dots, o_T)$, estimate model parameters $\lambda = (A, B, \Pi)$ that maximize $P(O|\lambda)$

3.1 Solution to Problem 1

Forward Algorithm [5]:

- Define forward variable $\alpha_t(i)$ as:

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = i | \lambda) \quad (1)$$

- $\alpha_t(i)$ is the probability of observing the partial sequence (o_1, o_2, \dots, o_t) such that the state q_t is i
- Induction

1. Initialization:

$$\alpha_1(i) = \pi_i b_i(o_1) \quad (2)$$

2. Induction:

$$\alpha_{t+1}(i) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \quad (3)$$

3. Termination:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (4)$$

Backward Algorithm [5]:

- Define backward variable $\beta_t(i)$ as:

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T, q_t = i | \lambda) \quad (5)$$

- $\beta_t(i)$ is the probability of observing the partial sequence (o_1, o_2, \dots, o_t) such that the state q_t is i

- Induction

1. Initialization:

$$\beta_T(i) = 1 \quad (6)$$

2. Induction:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1} \beta_{t+1}(j)), \quad (7)$$

$$1 \leq i \leq N, t = T - 1, \dots, 1$$

3. Termination:

$$P(O|\lambda) = \sum_{i=1}^N \beta_1(i) \quad (8)$$

3.2 Solution to Problem 2

Viterbi Algorithm [5]:

- Initialization:

$$\delta_1(i) = \pi_i b_i(o_1), 1 \leq i \leq N \quad (9)$$

$$1 \leq i \leq N$$

$$\psi_1 = 0 \quad (10)$$

- Recursion:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i)a_{ij}]b_j(o_t) \quad (11)$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i)a_{ij}]b_j(o_t) \quad (12)$$

$$1 \leq j \leq N, 2 \leq t \leq T$$

- Termination:

$$P^* = \max_{1 \leq i \leq N} [\delta_t(i)] \quad (13)$$

$$q_t^* = \arg \max_{1 \leq i \leq N} [\delta_t(i)] \quad (14)$$

- Backtracking:

$$q_t^* = \psi_t(q_{t+1}^*) \quad (15)$$

$$t = T-1, T-2, \dots, 1$$

3.3 Solution to Problem 3

Baum-Welch Algorithm [5]:

- Define $\xi(i, j)$ as the probability of being in state i at time t and in state j at time $t + 1$

$$\xi(i, j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{P(O|\lambda)} = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)} \quad (16)$$

- Define $\gamma_t(i)$ as the probability of being in state i at time t , given observation sequence.

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (17)$$

- $\sum_{t=1}^T \gamma_t(i)$ is the expected number of times state i is visited
- $\sum_{t=1}^T \xi_t(i, j)$ is the expected number of transition from state i to j

Update rules:

- $\bar{\pi}_i =$ expected frequency in state i at time ($t = 1$) = $\gamma_1(i)$
- $\bar{a}_{ij} =$ (expected number of transition from state i to state j)/(expected number of transitions from state i):

$$\bar{a}_{ij} = \frac{\sum_t \xi_t(i, j)}{\sum_t \gamma_t(i)} \quad (18)$$

- $\bar{b}_j(k) =$ (expected number of times in state j and observing symbol k)/(expected number of times in state j):

$$\bar{b}_j(k) = \frac{\sum_{t, o_t=k} \gamma_t(j)}{\sum_t \gamma_t(j)} \quad (19)$$

4 2D HMM

In paper [11], Yujian proposed definitions and proofs of 2D HMM. He has presented several analytic formulae for solving the three basic problems of 2-D HMM. Solution to Problem 2 is useful, and Viterbi algorithm can be easily adopted to image recognition with two dimensional input data. Unfortunately, solution to problem 1 and 3 may be used only with one dimensional data -observation vector. Besides presented solutions are for Markov model type "left-right", and not ergodic. So, I present solution to problems 1 and 3 for two dimensional data. The statistical parameters of the 2D model (Fig. 3):

- The number of states of the model N^2
- The number of data streams $k_1 \times k_2 = K$
- The number of symbols M
- The transition probabilities of the underlying Markov chain, $A = \{a_{ijl}\}$, $1 \leq i, j \leq N, 1 \leq l \leq N^2$, where a_{ijl} is the probability of transition from state ij to state l
- The observation probabilities, $B = \{b_{ijm}\}$, $1 \leq i, j \leq N, 1 \leq m \leq M$ which represents the probability of generating the m_{th} symbol in the ij_{th} state.
- The initial probability, $\Pi = \{\pi_{ijk}\}$, $1 \leq i, j \leq N, 1 \leq k \leq K$.
- Observation sequence $O = \{o_t\}$, $1 \leq t \leq T$, o_t is square matrix simply observation with size $k_1 \times k_2 = K$

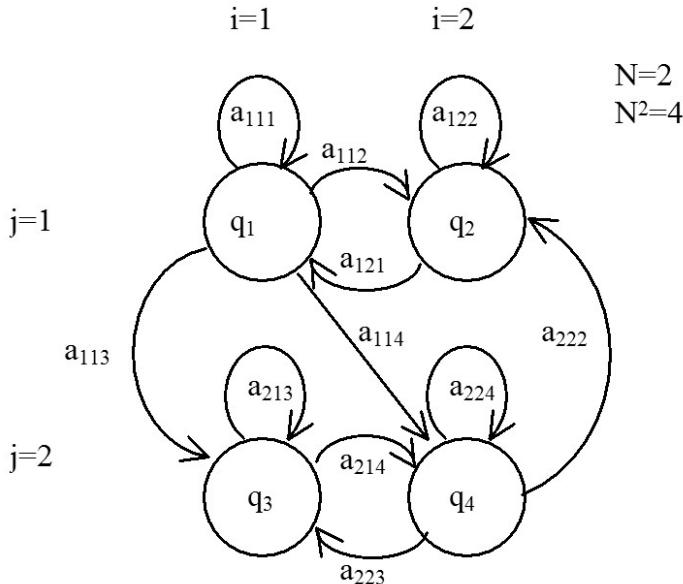


Fig. 3 Two-dimensional ergodic HMM

4.1 Solution to 2D Problem 1

Forward Algorithm

- Define forward variable \$\alpha_t(i, j, k)\$ as:

$$\alpha_t(i, j, k) = P(o_1, o_2, \dots, o_t, q_t = ij | \lambda) \quad (20)$$

- \$\alpha_t(i, j, k)\$ is the probability of observing the partial sequence \$(o_1, o_2, \dots, o_t)\$ such that the state \$q_t\$ is \$i, j\$ for each \$k_{th}\$ stream of data
- Induction

- Initialization:

$$\alpha_1(i, j, k) = \pi_{ijk} b_{ij}(o_1) \quad (21)$$

- Induction:

$$\alpha_{t+1}(i, j, k) = \left[\sum_{l=1}^N \alpha_t(i, j, k) a_{ijl} \right] b_{ij}(o_{t+1}) \quad (22)$$

- Termination:

$$P(O|\lambda) = \sum_{t=1}^T \sum_{k=1}^K \alpha_T(i, j, k) \quad (23)$$

4.2 Solution to 2D Problem 3

Parameters reestimation Algorithm:

- Define $\xi_t(i, j, l)$ as the probability of being in state ij at time t and in state l at time $t + 1$ for each k_{th} stream of data

$$\begin{aligned}\xi_t(i, j, l) &= \frac{\alpha_t(i, j, k) a_{ijl} b_{ij}(o_{t+1}) \beta_{t+1}(i, j, k)}{P(O|\lambda)} = \\ &= \frac{\alpha_t(i, j, k) a_{ijl} b_{ij}(o_{t+1}) \beta_{t+1}(i, j, k)}{\sum_{k=1}^K \sum_{l=1}^{N^2} \alpha_t(i, j, k) a_{ijl} b_{ij}(o_{t+1}) \beta_{t+1}(i, j, k)}\end{aligned}\quad (24)$$

- Define $\gamma_t(i, j)$ as the probability of being in state i, j at time t , given observation sequence.

$$\gamma_t(i, j) = \sum_{l=1}^{N^2} \xi_t(i, j, l) \quad (25)$$

- $\sum_{t=1}^T \gamma_t(i, j)$ is the expected number of times state ij is visited
- $\sum_{t=1}^{T-1} \xi_t(i, j, l)$ is the expected number of transition from state ij to l

Update rules:

- $\pi_{ijk} =$ expected frequency in state i, j at time ($t = 1$) = $\gamma_1(i, j)$
- $\bar{a}_{ij} =$ (expected number of transition from state i, j to state l)/(expected number of transitions from state i, j):

$$\bar{a}_{ijl} = \frac{\sum_t \xi_t(i, j, l)}{\sum_t \gamma_t(i, j)} \quad (26)$$

- $\bar{b}_{ij}(k) =$ (expected number of times in state j and observing symbol k)/(expected number of times in state j):

$$\bar{b}_{ij}(k) = \frac{\sum_{t, o_t=k} \gamma_t(i, j)}{\sum_t \gamma_t(i, j)} \quad (27)$$

5 Experimenting

The image database *Amsterdam Library of Object Images* was used in experimenting. It is a color image collection of one-thousand small objects, recorded for scientific purposes. In order to capture the sensory variation in object recordings, they systematically varied viewing angle, illumination angle, and illumination color for each object, and additionally captured wide-baseline stereo images. They recorded over a hundred images of each object, yielding a total of 110,250 images for the collection [1, 3].

In order to verify the method has been selected fifty objects. Three images for learning and three for testing have been chosen. The 2D HMM has been implemented with parameters $N = 5, N^2 = 25, K = 25, M = 50$. Wavelet transform has been chosen as features extraction technique. Table 1 presents results of experiments.

Table 1 Comparison of recognition rate

Method	Recognition rate [%]
Eigenvector	86
1D HMM	90
2D HMM	92

6 Conclusion

Article presents a new conception about two-dimensional hidden Markov models. We show solutions of principle problems for ergodic 2D HMM, which may be applied for 2D data. Recognition rate of the method is 92%, which is better than 1D HMM. Furthermore, the advantage of this approach is that there is no need to convert the input two-dimensional image on a one-dimensional data, what simplifies pattern recognition system.

References

1. Amsterdam Library of Object Images, <http://www.science.uva.nl/aloi>
2. Eickeler, S., Müller, S., Rigoll, G.: High Performance Face Recognition Using Pseudo 2-D Hidden Markov Models. In: European Control Conference (1999), <http://citeseer.ist.psu.edu>
3. Geusebroek, J.M., Burghouts, G.J., Smeulders, A.W.M.: The Amsterdam library of object images. Int. J. Comput. Vision 61(1), 103–112 (2005)
4. Joshi, D., Li, J., Wang, J.Z.: A computationally Efficient Approach to the estimation of two- and three-dimensional hidden Markov models. IEEE Transactions on Image Processing 15(7), 1871–1886 (2006)
5. Kanungo, T.: Hidden Markov Model Tutorial (1999), <http://www.kanungo.com/software/hmm tut.pdf>
6. Kubanek, M.: Automatic methods for determining the characteristic points in face image. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010, Part I. LNCS, vol. 6113, pp. 523–530. Springer, Heidelberg (2010)
7. Li, J., Najmi, A., Gray, R.M.: Image classification by a two dimensional Hidden Markov model. IEEE Transactions on Signal Processing 48, 517–533 (2000)
8. Rabiner, L.R.: A tutorial on hidden Markov models and selected application in speech recognition. Proc. IEEE 77, 257–285 (1989)
9. Samaria, F., Young, S.: HMM-based Architecture for Face Identification. Image and Vision Computing 12(8), 537–583 (1994)
10. Vitoantonio Bevilacqua, V., Cariello, L., Carro, G., Daleno, D., Mastronardi, G.: A face recognition system based on Pseudo 2D HMM applied to neural network coefficients. Soft Comput. 12(7), 615–621 (2008)
11. Yujian, L.: An analytic solution for estimating two-dimensional hidden Markov models. Applied Mathematics and Computation 185, 810–822 (2007)

Methods of Learning Classifier Competence Applied to the Dynamic Ensemble Selection

Maciej Krysmann and Marek Kurzynski

Abstract. In the design of dynamic ensemble selection system (DES) based on competence measure of base classifiers, two steps can be distinguished. The first step consists in calculating the set of competences of a classifier at all points of validation set. In the second step this set is generalized to the whole feature space or - in other words - the procedure of learning competence function using the competence set is performed. In the paper different methods of learning competence function are developed using the concept of randomized reference classifier as a basis for determining the competence set. The methods developed are: potential function method, linear regression, neural network, radial basis neural network, generalized regression neural network and 1-Nearest Neighbor method. Performance of DES systems for different learning methods were experimentally investigated using 9 benchmark data sets from the UCI Machine Repository and results of comparative analysis are presented.

1 Introduction

One of the most important tasks in optimizing a multiclassifier system (MCS) is the choice of a group of competent (adequate) classifiers from a pool of base classifiers [2, 7]. There are two main approaches to this problem: static selection scheme in which classifier (ensemble of classifiers) is selected for all test patterns, and dynamic selection method which explores the use of different classifiers for different test patterns.

For the calculation of the competence in the most dynamic ensemble selection (DES) systems, various performance (local accuracy) estimates are used, such as the

Maciej Krysmann · Marek Kurzynski

Department of Systems and Computer Networks, Wroclaw University of Technology,
Wyb. Wyspianskiego 26, 50-370 Wroclaw Poland

e-mail: {maciej.krysmann, marek.kurzynski}@pwr.wroc.pl

local accuracy estimation [4,9,12], Bayes confidence measure [6], multiple classifier behaviour [5], methods based on relating the response of the classifier with the response obtained by a random guessing [13] or randomized classification model [14], among others.

Regardless of the interpretation of the competence measure, the whole procedure of calculating the classifier competence can be divided into the following two steps [14]. In the first step, the set of competences at all points of validation set is calculated. In the second step, the competence measure (function) of the classifier is constructed. This construction is based on extending (generalizing) the competence set to the entire feature space. In other words, this step can be considered as a problem of learning competence function using the competence set.

In this study the second step of the above procedure is addressed by developing the different methods of learning competence functions with the probabilistic model and randomized reference classifier (RRC) [14] as a basis for determining the competence set. In the paper the following methods of learning competence functions are developed: potential function method, linear regression model, artificial neural network, radial basis neural network, generalized regression neural network and 1-Nearest Neighbor method.

The DES system based on the competence measure classifies a test object in the following manner. First, the competences are determined for each base classifier in the pool. Then a subset of the classifiers with the competences greater than the probability of random classification is selected from the pool for a test object. The selected classifiers are combined using the weighted majority voting rule with continuous-valued outputs, where the weights are equal to the competences. Finally, the DES system classifies x using the maximum rule. Performance of DES systems for different learning methods were experimentally investigated using 9 benchmark data sets from the UCI Machine Repository.

The paper is organized as follows. In section 3 the randomized reference classifier (RRC) is presented. Sections 3 describes the proposed methods of learning competence. Section 4 presents the DES system that was constructed using the competence model. The experiments conducted and results with discussion are presented in section 5. Section 6 concludes the paper.

2 Multiple Classifier System Based on Measure of Competence

In the multiclassifier (MC) system we assume that a set of trained classifiers $\Psi = \{\psi_1, \psi_2, \dots, \psi_L\}$ called base classifiers is given. A classifier ψ_l is a function $\psi_l : \mathcal{X} \rightarrow \mathcal{M}$ from a feature space to a set of class labels $\mathcal{M} = \{1, 2, \dots, M\}$. Classification is made according to the maximum rule

$$\psi_l(x) = i \Leftrightarrow d_{li}(x) = \max_{j \in \mathcal{M}} d_{lj}(x), \quad (1)$$

where $[d_{l1}(x), d_{l2}(x), \dots, d_{lM}(x)]$ is a vector of class supports (classifying function) produced by ψ_l . Without loss of generality we assume, that $d_{lj}(x) \geq 0$ and $\sum_j d_{lj}(x) = 1$.

The ensemble Ψ is used for classification through a combination function which, for example, can select a single classifier or a subset of classifiers from the ensemble, it can be independent or dependent on the feature vector x (in the latter case the function is said to be dynamic), and it can be non-trainable or trainable [7]. The proposed multiclassifier system uses dynamic ensemble selection (DES) strategy with trainable selection/fusion algorithm. The basis for dynamic selection of classifiers from the pool is a competence measure $c(\psi_l|x)$ of each base classifier ($l = 1, 2, \dots, L$), which evaluates the competence of classifier ψ_l i.e. its capability to correct activity (correct classification) at a point $x \in \mathcal{X}$. For the training of competence it is assumed that a validation set

$$\mathcal{V} = \{(x_1, j_1), (x_2, j_2), \dots, (x_N, j_N)\}; \quad x_k \in \mathcal{X}, \quad j_k \in \mathcal{M} \quad (2)$$

containing pairs of feature vectors and their corresponding class labels is available.

The construction of the competence measure consists of the two following steps. In the first step, a hypothetical classifier called a randomized reference classifier (RRC) is constructed. The RRC can be considered as equivalent to the classifier ψ_l and its probability of correct classification $Pc^{(RRC)}(x_k)$ can be used as the competence $C(\psi_l|x_k)$ of that classifier. In the second step, the competences $C(\psi_l|x_k)$, $x_k \in \mathcal{V}$ are used to construct the competence function $c(\psi_l|x)$. The construction is based on extending (generalizing) the competences $C(\psi_l|x_k)$ to the entire feature space \mathcal{X} . The next two sections describe the steps of the method in detail.

3 Randomized Reference Classifier (RRC)

The RRC is a stochastic classifier and therefore it is defined using a probability distribution over the set of class labels \mathcal{M} or, assuming the canonical model of classification, over the product of class supports $[0, 1]^M$. In other words, the RRC uses the maximum rule and a vector of class supports $[\delta_1(x), \delta_2(x), \dots, \delta_M(x)]$ for the classification of the feature vector x , where the j -th support is a realization of a random variable (rv) $\Delta_j(x)$. The probability distributions of the rvs $\Delta_j(x)$ are chosen in such a way that the following conditions are satisfied (throughout this description, the index l of the classifier ψ_l and its class supports is dropped for clarity):

- (1) $\Delta_j(x) \in [0, 1]$;
- (2) $E[\Delta_j(x)] = d_j(x)$, $j = 1, 2, \dots, M$;
- (3) $\sum_{j=1,2,\dots,M} \Delta_j(x) = 1$,

where E is the expected value operator. From the above definition it follows that the RRC can be considered as equivalent to the classifier ψ for the feature vector x since it produces, on average, the same vector of class supports as the modeled classifier.

Since the RRC performs classification in a stochastic manner, it is possible to calculate the probability of classification an object x to the i -th class:

$$P^{(RRC)}(i|x) = \Pr[\forall_{k=1,\dots,M, k \neq i} \Delta_i(x) > \Delta_k(x)]. \quad (3)$$

In particular, if the object x belongs to the i -th class, from (5) we simply get the conditional probability of correct classification $P_C^{(RRC)}(x)$.

The key element in the modeling presented above is the choice of probability distributions for the rvs $\Delta_j(x)$, $j \in \mathcal{M}$ so that the conditions 1-3 are satisfied. In this paper beta probability distributions are used with the parameters $\alpha_j(x)$ and $\beta_j(x)$ ($j \in \mathcal{M}$). The justification of the choice of the beta distribution, resulting from the theory of order statistics can be found in [14].

Applying the RRC to a validation point x_k and putting in (5) $i = j_k$, we get the probability of correct classification of RRC at a point $x_k \in V$:

$$\begin{aligned} P_C^{(RRC)}(x_k) &= \int_0^1 b(u, \alpha_1(x_k), \beta_1(x_k)) \\ &\quad [\prod_{j=2}^M B(u, \alpha_j(x_k), \beta_j(x_k))] du, \end{aligned} \quad (4)$$

where $B(\cdot)$ is a beta cumulative distribution function. The MATLAB code for calculating probabilities (4) was developed and it is freely available for download.

Since the RRC can be considered equivalent to the modeled base classifier $\psi_l \in \Psi$, it is justified to use the probability (4) as the competence of the classifier ψ_l at the validation point $x_k \in \mathcal{V}$, i.e.

$$C(\psi_l|x_k) = P_C^{(RRC)}(x_k). \quad (5)$$

4 Learning Classifier Competence

4.1 Preliminaries

After the first step of the method the competence set is given for each base classifier ψ_l ($l = 1, 2, \dots, L$) of an ensemble Ψ :

$$\mathcal{C}_l = \{(x_1, C(\psi_l|x_1)), (x_2, C(\psi_l|x_2)), \dots, (x_N, C(\psi_l|x_N))\}. \quad (6)$$

In the second step, based on the competence set (6), the competence function c_l is determined. In other words, information contained in the set \mathcal{C}_l , i.e. values of competence for validation points $x_k \in \mathcal{V}$, is generalized to the whole feature space \mathcal{X} . This means that we are dealing with the problem of learning the competence function, to solve which different methods of machine learning can be applied. In the

next part a set of such methods is presented, which differ in properties, calculation procedure and interpretation of the set (6).

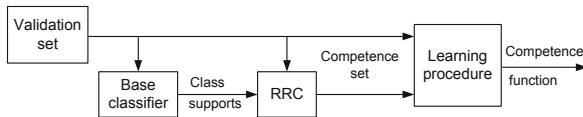


Fig. 1 Flowchart of calculating the competence function of a base classifier

4.2 Methods of Learning Classifier Competence

4.2.1 Potential Function

Method approximates competence at a point x_m from potential function, which for each calculation use all validation points $x_k \in \mathcal{V}$. It can be easily seen (7) that method is based on distances from point x_m to each point from \mathcal{V} . That approximation method doesn't require learning phase so it is instantly ready to work (after the \mathcal{V} set is created). Disadvantage of potential function is requirement of storing in memory whole \mathcal{V} set.

$$C(x_m) = \frac{\sum_{k=1}^N C_k * K(x_m, x_k)}{\max(C_k)}. \quad (7)$$

where

$$K(x_m, x_k) = e^{-d(x_m, x_k)}. \quad (8)$$

In this method an improvement can be achieved by bootstrapping used on set \mathcal{C}_l . Bootstrapping procedure generates new 2-times bigger set using \max function from \mathcal{C}_l . Test presented gain of quality form 0.16% - 3%.

4.2.2 Regression Method

Method assumes multidimensional linear dependency between features vector and value of competence. Let object x be described as follows: $x = [x_1, x_2, \dots, x_g]$, where g is amount of features in particular classification task (index m is dropped for clarity).

$$C(x) = Ax + a_0. \quad (9)$$

where

$$A^T = [a_1, \dots, a_g]. \quad (10)$$

Higher degree regression methods were rejected because of computational complexity and limited size of the validation set - for most cases (even for second degree)

there are too much parameters to estimate from the validation set, which results in random work of algorithm.

4.2.3 Neural Network Method

Several configurations of neural networks (different numbers of neurons in hidden layers, sigmoid transfer function) were tested to choose best parameters configuration. Those parameters were used in further discussed comparison tests. As input of network was used features vector, which allowed network to approximate value of competence on output. Finally used network was back propagation trained network with 10 neurons in each of 2 hidden layers and one sigmoid neuron in output layer.

4.2.4 Radial Basis Neural Network

Specialized type of a neural network which have a radial transfer function [1] (11) - argument (n) is value of adder in each of neurons of radial basis layer. That layer is one of two in network - the second one is typical linear layer.

$$rad(n) = e^{-n^2} \quad (11)$$

Network input vectors consist of distances between new object (x_m) and objects from learning set. Output is a single value calculated in linear layer, which weights are trained assuming linear dependence between target vector and simulated inputs for this layer. Learning process, which also creates structure of network, is limited by achieving desired mean square error (0.05).

4.2.5 Generalized Regression Neural Network

Generalized regression neural network [10] is a special type of the radial basis network. First layer is exactly the same as in the radial basis network (with the same transfer function 11), only difference is in linear layer - weights are set to target vectors, without having linear relation "target - linear layer inputs". As in the radial basis neural network, learning (and building network structure) process ends when mean square error reaches 0.05.

4.2.6 1-Nearest Neighbor Method

Method based on idea of nearest neighbor classifier. Competence for the new point x_m is weighted value of competence from point ($x_k \in \mathcal{V}$) which have the a smallest distance to x_m . As weight was assumed the distance x_k to x_m in order to account precision level of this approximation.

$$C(x_m) = \frac{C(x_k : \min(d(x_m, x_k)))}{d(x_m, x_k)} \quad (12)$$

5 Dynamic Ensemble Selection Systems

The probabilistic model of classifier competence is a basis for construction of dynamic ensemble selection (DES) system. This system classifies the test object $x \in \mathcal{X}$ in the following manner. First, the subset of classifiers $\Psi(x)$ from the pool of base classifiers is selected, which competence values are greater than the probability of random classification. Next, the selected classifiers are combined on continuous-valued level, i.e. for a j th class the support is calculated as a weighted sum of supports where the weights are equal to the competence values, viz.

$$d_j(x) = \sum_{\psi_t \in \Psi(x)} C(\psi_t | x) d_{tj}(x) \quad (13)$$

and the final classification is made according to the maximum rule. In the case when subset $\Psi(x)$ is empty, the test object is classified by random guessing.

6 Experiments

6.1 Experiment Setup

Several Experiments were made, in order to evaluate a performance of methods of learning classifier competence. Experiments were performed in MATLAB with own procedures and PRTools toolbox [8] for base classifiers implementation. Benchmark databases (Table 1) used in tests were acquired from UCI Machine Learning repository [11].

Table 1 Databases used in tests

Database Objects Features Classes			Database Objects Features Classes				
Dermathology	366	34	6	Iris	150	4	3
Ecoli	336	7	8	Pima Diabetes	768	8	2
Glass	214	9	6	Wine	178	13	3
Haberman	306	3	2	Yeast	1484	8	10
Ionosphere	351	34	2				

To obtain results two-fold cross-validation methodology was used. There were tested 6 methods, using 9 databases. Pool used in experiments was builded from following base classifiers (the same as in previous work [14]):

- k-nearest neighbors classifiers ($k=1,5,15$)
- nearest mean classifier
- Parzen density classifier (Gaussian kernel with h_{opt} and $\frac{h_{opt}}{2}$)
- decision tree with Gini splitting criterion

- neural network classifier (two hidden layers with 5 neurons each and one hidden layer with 10 neurons, both maximum 80 epochs of training).

6.2 Results and Discussion

Table 2 presents results of experiments. Values are percent of correctly classified objects for each database and method of learning competence. Last row presents average rank for each method. Last column presents results from bootstrapped potential function for 2-class cases. Average rank for this column is calculated using bootstrapped (only 2-class problems) and “classical” potential function results (for the rest). Because bootstrapped modification improves results only in 2-class cases there is no statistical difference comparing it to non-bootstrapped method in general.

To compare general quality (percent of correct answers) and measure differences between methods, there were used statistical tests [3]. Firstly the Friedman test (with Iman-Davenport correction) was performed to determine if all the methods have equal quality on statistically significant level. This test indicated that there are differences between the methods and in order to reveal those differences there was performed post-hoc Holm step-down procedure. Results from this statistical test are presented in a Table 3, as follows: H_0 means that two methods are similar in

Table 2 Experimental results

Database	RB	NN	GR	NN	1-NN	NN	lin. reg	Pot. func.	Bstr.	P.f.
EColi	75.59	84.83	74.23	83.81	84.73	84.61	—	—	—	—
Ionosphere	84.57	85.06	68.42	84.20	86.37	84.12	84.28	—	—	—
Iris	95.78	95.71	80.47	94.90	95.77	95.57	—	—	—	—
Wine	71.07	44.07	33.43	72.43	76.44	74.80	—	—	—	—
Pima	67.63	41.77	34.56	70.89	67.71	68.37	72.28	—	—	—
Glass	79.39	84.48	53.85	82.01	83.24	82.56	—	—	—	—
Haberman	73.74	72.53	73.07	74.56	74.16	74.37	—	—	—	—
Derma	96.49	96.59	61.45	96.45	96.73	96.87	97.36	—	—	—
Yeast	46.98	56.31	46.12	55.77	56.45	55.37	—	—	—	—
Avg. rank	3.89	3.11	5.89	3.33	2.00	2.89	(2.67)	—	—	—

Table 3 Accepted hypothesis table

	RB	NN	GR	NN	1-NN	NN	lin. reg	Pot. func.	
RB NN	X	H_0	H_1	H_0	H_1	H_0			
GR NN	H_0	X	H_1	H_0	H_0	H_0			
1-NN	H_1	H_1	X	H_1	H_1	H_1			
NN	H_0	H_0	H_1	X	H_0	H_0			
lin. reg	H_1	H_0	H_1	H_0	X	H_0			
Pot. func.	H_0	H_0	H_1	H_0	H_0	X			

statistical way (hypothesis H_0 is accepted), H_1 means that they are different and $\neg H_0$ - test couldn't reject H_0 hypothesis. It can be easily seen that the 1-NN method is worse than the others, because of really naive assumption of working. However for other methods Holm's procedure accepts null-hypothesis H_0 on significance level $p = 0.05$, what means that those ways of learning classifier competence are similar in general (except radial basis neural network, which is worse than linear regression, but similar to others). However linear regression have lowest standard deviation in ranks - that indicates the best stability, which combined with the lowest rank (best results) implies that the linear regression is a safest (but not always best) choice.

7 Final Remarks

Paper describes research into methods of learning (approximation) classifier competence from competence set. Several methods were tested on well known benchmark databases. Results have proven that in general there are no statistical differences in quality (measured for all cases) between methods. However linear regression and potential function, considering particular databases (not general performance), were slightly better than the others. These two methods also are the less expensive in implementation and learning time. Therefore linear regression and potential function should be checked at first while searching for optimum method for a real-life problem. Other methods however can not be rejected and should also be considered in a real task.

General results presented in paper proven that choice of the learning classifier competence method, does not have influence on this version of DES system. However more sophisticated methods such as neural networks could be not fully utilized and that justifies further development of the Dynamic Ensemble Selection with RRC systems taking into account presented different methods of learning classifier competence.

References

1. Buhmann, M.D.: Radial Basis Functions: Theory and Implementations. Cambridge University (2003)
2. Canuto, A., Abreu, M., Oliveira, L.: Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection and fusion methods for ensembles. Pattern Recognition Letters 28, 472–486 (2007)
3. Demsar, J.: Statistical comparison of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30 (2006)
4. Didaci, L., Giacinto, G., Roli, F., Marcialis, G.: A study on the performance of dynamic classifier selection based on local accuracy estimation. Pattern Recognition 38, 2188–2191 (2005)
5. Giacinto, G., Roli, F.: Dynamic classifier selection based on multiple classifier behaviour. Pattern Recognition 34, 1879–1881 (2001)

6. Huenupan, F., Yoma, N., Molina, C., Garreton, C.: Confidence based multiple classifier fusion in speaker verification. *Pattern Recognition Letters* 29, 957–966 (2008)
7. Kuncheva, L.: Combining Pattern Classifiers: Methods and Algorithms. Wiley Inter-science, New Jersey (2004)
8. PRTools (2012), <http://www.prtools.org/>
9. Smits, P.C.: Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection. *IEEE Trans. on Geoscience and Remote Sensing* 40(4), 801–813 (2002)
10. Specht, D.F.: A general regression neural network. *IEEE Trans. Neural Networks* 2, 568–576 (1991)
11. UCI Machine Learning repository, <http://archive.ics.uci.edu/ml/>
12. Woods, K., Kegelmeyer, W., Bowyer, K.: Combination of multiple classifiers using local accuracy estimates. *IEEE TRAns. on Pattern Analysis and Machine Learning* 19(4), 405–410 (1997)
13. Woloszynski, T., Kurzynski, M.: On a new measure of classifier competence in the feature space. In: Kurzynski, M., Wozniak, M. (eds.) Computer Recognition Systems 3. AISC, vol. 57, pp. 285–292. Springer, Heidelberg (2009)
14. Woloszynski, T., Kurzynski, M.: A probabilistic model of classifier competence for dynamic ensemble selection. *Pattern Recognition* 44, 2656–2668 (2011)

The Method of Improving the Structure of the Decision Tree Given by the Experts

Robert Burduk

Abstract. This paper presents the problem of sequential decision making in the pattern recognition task. This task can be presented using a decision tree. In this case, it is assumed that the structure of the decision tree is determined by experts. The classification process is made in each node of the tree. This paper proposes a way to change the structure of the decision tree to improve the quality of classification. The split criterion is based on the confusion matrix. The obtained results were verified on the basis of the example of the computer-aided medical diagnosis.

1 Introduction

There are many alternatives to represent the classification task. One of them is a multistage (sequential) approach that allows to break up a complex decision into a collection of several simpler decisions [9], [13]. The multistage pattern recognition has two approaches: a decision tree classifier and a hierarchical classifier [10]. The decision tree approach built the classification model and their tree structure is construct in the learning process [11], [12]. Each nonleaf node of the decision tree is associated with a decision function.

The synthesis of the hierarchical classifier is a complex problem. It involves a specification of the following components [5], [13]:

- design of a decision tree structure,
- selection of features used at each non-terminal node of the decision tree,
- the choice of decision rules for performing the classification.

In many approaches the decision tree structure is fixed before the learning process [8]. In this work, this case will be the basis for further discussion. In particular,

Robert Burduk

Department of Systems and Computer Networks, Wroclaw University of Technology,
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
e-mail: robert.burduk@pwr.wroc.pl

this paper proposes a way to change the structure of the decision tree to improve the quality of classification. The split criterion is based on the confusion matrix.

The content of the work is as follows. Section 2 introduces the idea of the hierarchical classifier. In the next section we describe a split criterion based on the confusion matrix. In Section 3 we present the results of experimental research on the medical diagnosis. The last section concludes the paper.

2 Hierarchical Classifier

The hierarchical classifier contains a sequence of actions [2], [7]. These actions are simple classification tasks executed in the individual nodes of the decision tree. Some specific features are measured on every nonleaf node of the decision tree. At the first nonleaf node features x_0 are measured, at the second features x_1 are considered and so on. Every set of features comes from the whole vector of features. In every node of the decision tree the classification is executed according to the specific rule. The decisions i_0, i_1, \dots, i_N are the results of recognition in the suitable node of the tree. The logic of making the decision is represented using the decision tree. The design of a decision tree structure in our approach to the hierarchical classifier is based on human expert knowledge. In our task of classification the number of classes is equal to NC . The terminal nodes are labeled with the number of the classes from the $M = 1, 2, \dots, NC$, where M is the set of labels classes. The non-terminal nodes are labeled by numbers of 0, $NC+1$, $NC+2$ reserving 0 for the root-node. Let us introduce the notation for the received model of multistage recognition [5]:

- $\overline{\mathcal{M}}$ – the set of internal (nonleaf) nodes,
- \mathcal{M}_i – the set of class labels attainable from the i -th node ($i \in \overline{\mathcal{M}}$),
- \mathcal{M}^i – the set of nodes of immediate descendant node i ($i \in \overline{\mathcal{M}}$),
- m_i – node of direct predecessor of the i -th node ($i \neq 0$).

In each interior node the decision rule (recognition algorithm) is used. It maps observation subspace to the set of the immediate descendant nodes of i -th node [1], [4]:

$$\Psi_i : X_i \rightarrow \mathcal{M}^i, \quad i \in \overline{\mathcal{M}}. \quad (1)$$

This approach minimizes the misclassification rate for the particular nodes of a tree. The decision rules at each node are mutually independent. There are no relationships between the nodes.

3 Split Criteria

Now we propose the split criterion. The division of the internal node will be made on the basis of the multidimensional confusion matrix. Specifically, we propose the division associated with an incorrect classification. This division is binary, which

means that the node that will meet the criterion of the split will have two child nodes. One of them represents a new internal decision tree node. The second one represents the label of a class that met the appropriate condition. The internal node of the decision tree is analysed in detail via the multidimensional confusion matrix. The columns of the confusion matrix correspond to the predicted labels (decisions made by the classifier in the internal node). The rows correspond to the true class labels. In this matrix diagonal elements represent the overall performance of each label. The off-diagonal elements represent the errors related to each label.

Now we present the split criterion. For every successor of the internal node $i \in \mathcal{M}$ we create the $L \times L$, $L = \text{card}(\mathcal{M}^i)$ dimensional confusion matrix. Now we calculate the set of factors $W(k_l^i)$, where $l = 1, 2, \dots, L$ is the number of successors of node i , according to the formula:

$$W(k_l^i) = \sum_{m=1, m \neq l}^L w_{l,m}^i + \sum_{m=1, m \neq l}^L w_{m,l}^i. \quad (2)$$

The example of the confusion matrix is presented in Tab. 1.

Table 1 The confusion matrix for the nonleaf node i

		estimated			
		k_1^i	k_2^i	\dots	k_L^i
true	k_1^i	$w_{1,1}^i$	$w_{1,2}^i$	\dots	$w_{1,L}^i$
	k_2^i	$w_{2,1}^i$	$w_{2,2}^i$	\dots	$w_{2,L}^i$
\vdots	\vdots	\vdots	\vdots	\dots	\vdots
k_L^i	$w_{L,1}^i$	$w_{L,2}^i$	\dots	$w_{L,L}^i$	

Then we are looking for the lowest and highest values among all $W(k_l^i)$ for fixed i . If the absolute value of the difference exceeds 0.5, we split node i . If this condition is not fulfilled we do not split node i . The structure of the decision tree is therefore unchanged.

4 Experiments

The first mathematical model of acute abdominal pain (APP) with the decision tree was given in [6]. This model has sixteen classes and four stages of recognition. The expert physicians (from the Surgical Clinic Wroclaw Medical Academy) provided the decision tree presented in Fig.1. It has eight classes and two stages. The labels present the following disease units of the AAP:

1. cholecystitis,
2. pancreatitis,

3. non-specific abdominal pain,
4. rare disorders of acute abdominal,
5. appendicitis,
6. diverticulitis,
7. small-bowel obstruction,
8. perforated peptic ulcer.

Numbers of leaves are the number of the diagnosis and the numbers in the nodes correspond to the following diagnoses:

1. acute enteropathy,
2. acute disorders of the digestive system,
3. others.

The dataset used in the experiment consists of 476 clinical histories. This set is represented by 31 features. The selection of features has been made in accordance with the suggestions of another work on this topic [3], [6].

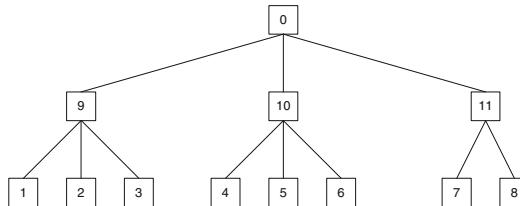


Fig. 1 Decision tree given by the experts

In our experiment, we used the k-NN rule and neutral networks (NN) in each internal node of the decision tree. K-NN classifier was tested for $k = 3, 5, 7$. In case with the neutral networks we used multilayer perceptron which has one hidden layer, with a linear combination functions in the hidden and output layers and a sigmoid activation functions in the hidden layers. The experiments were carried out for the number of neurons $n = 9, 12$. The experiments were carried out in SAS Enterprise Miner 6.1 environment [14]. The Tab. 2 presents the results of the classification for all internal nodes. The classification error was calculated according to 10-fold cross-validation method.

Table 2 The error in each internal nodes - decision tree given by the experts

Node	Classifier				
	3-NN	5-NN	7-NN	NN-9	NN-12
0	0.161	0.193	0.173	0.115	0.065
9	0.35	0.5	0.35	0.25	0.2
10	0.173	0.239	0.282	0.195	0.152
11	0.0	0.048	0.048	0.0	0.0

After the experiments $W(k_i^j)$ factors were calculated for each internal node. Their values are shown in Tab. 3–5. The line with the adopted rule split changed the structure of the decision tree. The proposed structures are shown in Fig. 2, and the change relates to the nodes labeled as 10 and 9 from the original decision tree. For node 9 two possible divisions were tested. One of them is the result of the k-NN classifier used in this node (version 1). The second one was proposed by neural network classifiers (version 2).

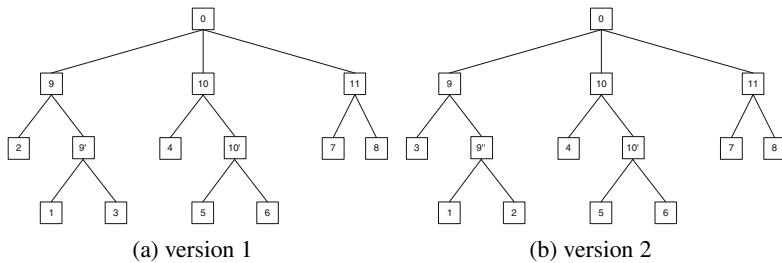


Fig. 2 The proposed changes in the structure of the decision tree

After changing the structure of the decision tree we repeated the experiments, and the results are shown in Tab. 6 and 7.

Table 3 The factors for the split criteria in node 0

Classifier	Factor for node		
	W(9)	W(10)	W(11)
3-NN	0.805	0.532	0.662
5-NN	0.717	0.565	0.717
7-NN	0.821	0.523	0.683
NN-9	0.909	0.473	0.618
NN-12	0.774	0.548	0.677

The results presented in this work suggest that you can still improve the quality of classification of a medical diagnosis for this problem. Improving the quality of the classification is related to the change in the original structure of the decision tree proposed by experts. This change is made on the basis of the split criterion proposed in this work. The error in this case is measured as the quality of the classification in each internal node. When the structure is changed, the error is the sum of errors from the respective nodes. The better quality of classification after the division of the node is marked as bold in Tab. 6 and 7.

Table 4 The factors for the split criteria in node 9

Classifier	Factor for node		
	W(1)	W(2)	W(3)
3-NN	1	0.273	0.727
5-NN	1	0.333	0.667
7-NN	1	0.417	0.583
NN-9	1	0.8	0.2
NN-12	1	0.833	0.167

Table 5 The factors for the split criteria in node 10

Classifier	Factor for node		
	W(4)	W(5)	W(6)
3-NN	0.2	0.8	1
5-NN	0.4	0.6	1
7-NN	1	0.417	0.583
NN-9	0	1	1
NN-12	0	1	1

Table 6 The error in each internal nodes - changed the structure of the tree - version 1

Node	Classifier				
	3-NN	5-NN	7-NN	NN-9	NN-12
9	0.111	0.066	0.066	0.155	0.088
9'	0.142	0.166	0.119	0.095	0.047
9+9'	0.253	0.232	0.185	0.25	0.135
10	0.0	0.0	0.105	0.105	0.052
10'	0.266	0.266	0.266	0.133	0.2
10+10'	0.266	0.266	0.371	0.238	0.252

Table 7 The error in each internal nodes - changed the structure of the tree - version 2

Node	Classifier				
	3-NN	5-NN	7-NN	NN-9	NN-12
9	0.066	0.088	0.066	0.044	0.044
9''	0.135	0.081	0.081	0.162	0.135
9+9''	0.201	0.169	0.147	0.206	0.179
10	0.0	0.0	0.105	0.105	0.052
10'	0.266	0.266	0.266	0.133	0.2
10+10'	0.266	0.266	0.371	0.238	0.252

5 Conclusions

The paper proposes a split criterion for the internal node of the decision tree which structure is given by experts. The criteria is related to the quality of the classification in an internal node. In order to calculate $W(k_j^i)$ coefficients the confusion matrix is used. The values of these coefficients are based on the incorrect classification of the analyzed internal node of the decision tree. With a fulfilled criteria a binary split of analyzed decision node is carried out. The new structure of the tree is proposed to improve the quality of classification. Experiments done in the work show that we obtain improved quality of the classification.

In the work before and after the change of the structure in all internal nodes the same classifier was used. Therefore, further research may relate to the use of an ensemble of a classifier in each internal node of the decision tree. Future work may also apply to the selection of classifiers for specific internal nodes of the decision tree.

Acknowledgements. This work is supported by the Polish National Science Center under a grant N N519 650440 for the period 2011-2014.

References

1. Berger, J.: Statistical Decision Theory and Bayesian Analysis. Springer, New York (1993)
2. Burduk, R.: Classification error in Bayes multistage recognition task with fuzzy observations. *Pattern Analysis and Applications* 13(1), 85–91 (2010)
3. Burduk, R., Woźniak, M.: Different decision tree induction strategies for a medical decision problem. *Central European Journal of Medicine* 7(2), 183–193 (2010)
4. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley and Sons (2000)
5. Kurzyński, M.: Decision Rules for a Hierarchical Classifier. *Pat. Rec. Let.* 1, 305–310 (1983)
6. Kurzyński, M.: Diagnosis of acute abdominal pain using three-stage classifier. *Computers in Biology and Medicine* 17(1), 19–27 (1987)
7. Kurzyński, M.: On the Multistage Bayes Classifier. *Pattern Recognition* 21, 355–365 (1988)
8. Manwani, N., Sastry, P.S.: Geometric decision tree. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 42(1), 181–192 (2012)
9. Mitchell, T.M.: Machine Learning. McGraw-Hill Comp., Inc., New York (1997)
10. Mui, J., Fu, K.S.: Automated classification of nucleated blood cells using a binary tree classifier. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-2*, 429–443 (1980)
11. Penar, W., Woźniak, M.: Experiments on classifiers obtained via decision tree induction methods with different attribute acquisition cost limit. In: Kurzynski, M., et al. (eds.) Computer Recognition Systems. ASC, vol. 45, pp. 371–377. Springer, Heidelberg (2007)
12. Quinlan, J.R.: Induction on Decision Tree. *Machine Learning* 1, 81–106 (1986)
13. Safavian, S.R., Landgrebe, D.: A survey of decision tree classifier methodology. *IEEE Trans. Systems, Man Cyber.* 21(3), 660–674 (1991)
14. Getting Started with SAS Enterprise Miner 6.1,
<http://support.sas.com/documentation/onlinedoc/miner>

Part II

Biometrics

Face Detection and Recognition under Heterogeneous Database Based on Fusion of Catadioptric and PTZ Vision Sensors

Aditya Raj, Redouane Khemmar, Jean Yves Eratud, and Xavier Savatier

Abstract. Large field of view with high resolution has always been sought-after for Mobile Robotic Authentication. So the vision system proposed here is composed of a catadioptric sensor for full range monitoring and a Pan Tilt Zoom (PTZ) camera together forming an innovative sensor, able to detect and track any moving objects at a higher zoom level. In our application, the catadioptric sensor is calibrated and used to detect and track Regions Of Interest (ROIs) within its 360 degree Field Of View (FOV), especially face regions. Using a joint calibration strategy, the PTZ camera parameters are automatically adjusted by the system in order to detect and track the face ROI within a higher resolution and project the same in facespace for recognition via Eigenface algorithm. The whole development has been partially validated by application for the Face recognition using our own database.

1 Introduction

Ample amount of development has been released with multiple-camera sensor systems to meet the rapidly growing demands in monitoring mobile robot applications. One of the recent examples is Nomad Biometric Authentication (NOBA¹) system

Aditya Raj

Research Institute for Embedded Systems (IRSEEM), ESIGELEC, 76801 Saint Etienne du Rouvray Cedex, Rouen, FRANCE and Vellore Institute of Technology (VIT University), Vellore, India

e-mail: luvableaditya@gmail.com

Aditya Raj · Redouane Khemmar · Jean Yves Eratud · Xavier Savatier

IRSEEM/ESIGELEC, 76801 Saint Etienne du Rouvray Cedex, Rouen, France

e-mail: {khemmar, ertaud, savatier}@esigelec.fr

¹ This work is part of the Nomad Biometric Authentication (NOBA) project funded by ERDF under the Interreg IVA program (Ref. No. 4051) in collaboration with the University of Kent.

that considers the field of biometrics and mobile robotics which are currently dissociated despite being based on common technological foundations (perception, detection and classification). A unique example is the use of an omnidirectional camera in combination with PTZ camera, referred to as a dual camera system (as shown in Fig.1). Omnidirectional cameras are able to exploit a wide Field Of View (FOV) within its 360 degrees for full range monitoring. However, low and non-uniform resolution of these catadioptric sensors make close observations of particular targets, especially in biometric authentication applications difficult. PTZ cameras with high mobility and zoom ability, compensates the deficiencies of omnidirectional cameras. Based on a unified model projection introduced by Geyer [6], the catadioptric sensor is calibrated, in order to generate correct perspective images. By using Viola and Jones algorithm [12] in the resulting images, the program detects the face ROI, and then applies the tracking algorithm based on a correlation approach. Then a joint calibration method is performed to localize the face ROI in order to generate a zoomed in face image with high resolution. Finally, the extracted zoomed face is projected into face space for face recognition using Eigenface algorithm. This paper is organized as follows. Section 2 describes the system architecture and processing, modeling and the calibration of catadioptric sensor. Section 3 focuses on the proposed face detection and tracking algorithms. The fusion procedure of the omnidirectional and PTZ cameras is described in section 4. Section 5 allows to describe the implementation of Eigenface algorithm on our small database. Section 6 illustrates our experiment results, and section 7 concludes this paper.



Fig. 1 Prototype of Vision System

2 System Architecture and Processing

In the presented work, we focus on the issues related to face detection, extraction and recognition in a two camera serial architecture. The developed system is characterized by unbalanced sensors functionalities : A data exchange program from

the processing unit is designed to detect the domain of interest through the wide-field catadioptric sensor. The main program is performed to manage the processing and the pointing of the PTZ camera into the corresponding region of interest. Commanding and controlling position of the PTZ camera is established by using a network connection through http sockets (virtual channels), where messages and control information are sent and received.

2.1 The Catadioptric Sensor

Fixed View Point Constraint

The architecture of catadioptric sensor adheres to the Single-View-Point theory [1]. The SVP constraint enables to generate correct perspective images. In fact, the optical center of the camera has to coincide with the second focus F of the hyperbola located at distance $2e$ from the mirror focus as illustrated in Fig.2. The eccentricity " e " is a parameter of the mirror given by the manufacturer. To carry out this task, we first calibrate our camera with a standard calibration tool to determine the central point and the focal length. Knowing the parameters of both the mirror and the camera, the image of the mirror on the image plane can be easily predicted if the SVP constraint is taken into consideration, as shown in Fig.2. The expected mirror boundaries are superposed on the image and the mirror has then to be moved manually to fit this estimation as shown in Fig.3.

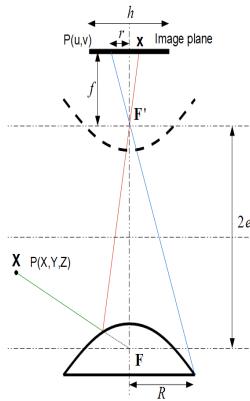


Fig. 2 Image Formation with an Hyperbolic Mirror

2.2 Sensor Calibration

Sensor calibration is the process to determine the optical and geometrical features which are generally addressed as intrinsic and extrinsic parameters and they allow to estimate the correspondence between 3D points of the scene and their projection into the image plane (pixel coordinates). The camera calibration we use, is based on a generic model introduced by [6] and [2], and then, modified by [9], who generalized the projection matrix and to take distortions into consideration. Further calibration details can be found in [4] and [10]. Figure 5 shows the projection process.

As described in [2] and [9] as well as mentioned in [7], the projection of a 3D point can be done by projecting the 3D point $X[w\ y\ z]$ onto the unit sphere centered on C_m : then Points $X_s = [x_s\ y_s\ z_s]^T$ are, then, projected onto the new frame with the origin $C_p = [0\ 0\ \xi]^T$, the obtained point $(X_s)_{C_p}$ onto a normalized plane and finally the last step enables us to find the camera projection matrix K expressed according to γ_u and γ_v , which are respectively, the generalized horizontal and vertical focal length, and (u_0, v_0) the coordinates of the principal point on the image point and the skew α :

$$p = K \cdot m = \begin{pmatrix} \gamma_u & \gamma_u \cdot \alpha & u_0 \\ 0 & \gamma_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot m \quad (1)$$

In our model, we consider that the impact of the parameter α , often null, is irrelevant. Parameters to be estimated in that model are: ξ , γ_u , γ_v , u_0 and v_0 .

With the tool developed by [3], calibration is achieved by observing a planar pattern at different positions. The pattern can be freely moved (the motion does not need to be known) and the user needs to select the four points corners pattern. This calibration process is similar to that of Mei [9]. It consists of a minimization over all the model parameters of an error function between the estimated projections of the pattern corners and the measured projection using Levenberg-Marquardt algorithm [8]. This minimization process enables to find the parameters combination that reduces the error of pattern retro projection.

3 Face Detection and Tracking-Based on Merger of Catadioptric Sensor and PTZ Camera

Most of image processing techniques are performed on conventional images, i.e. perspective images. Actually, deformations caused by the catadioptric system does not give us the opportunity to perform the existing face detection algorithms on raw images (Fig 4.). As a consequence, geometrical transformations should be performed to obtain a panoramic image close to perspective images, where the face detection algorithm will be performed.

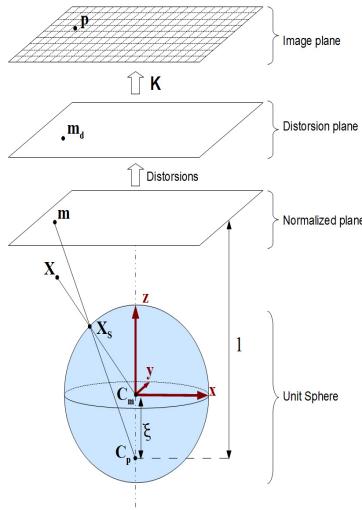


Fig. 3 Unified Projection Model

3.1 Panoramic Images Unwrapping

The existence of the model of the unit sphere, simplifies the unwrapping problem. Under the fixed view point constraint and by performing a retro projection, we are able to project the pixels of the panoramic pictures onto the unit sphere used in the unified model. Then, these pixels are projected onto the image plane. Thereby, we obtain the mapping between the pixels on the panoramic image and their corresponding on the camera retinal plane. Figure 5 shows the result of the spherical transformation of the considered catadioptric image. Face detection algorithm is applied to these unwrapped images.

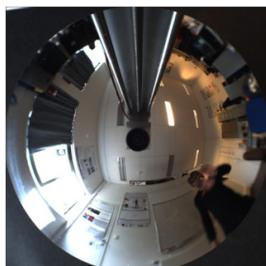


Fig. 4 Original catadioptric image



Fig. 5 Unwrapped image obtained by spherical transformation

3.2 Face Detection and Tracking Algorithms

Face detection techniques have been researched for years and much progress has been proposed in literature. However, in 2001, Paul Viola and Michael Jones [12] achieved a robust real time method for face detection, which was fifteen times quicker than the methods existing at that time. The technique relies on the use of simple Haar-like features that are evaluated quickly through the use of a new image representation called “integral image” that allows fast feature evaluation. Figure 6 shows two of the 60000 Haar-like features available.

To select the best filtering feature, Adaboost, the machine learning introduced in [5], is used. In fact, given a set of weak classifiers, not much better than random, if we iteratively combine their output, the training error will quickly converge to zero. Figure 7.(a) & Figure 7.(b) shows the detected and zoomed face image. As explained in [7] the face ROI so obtained is zoomed in with a high resolution which further can be processed for face recognition.



Fig. 6 Haar Like Features

4 Merger of Omnidirectional and PTZ Cameras

4.1 Joint Calibration Strategy

The joint calibration method is based on defining a reference position (x_0, y_0) for the PTZ camera on the catadioptric 360° image. This position has to coincide with the default orientation of the PTZ camera and is chosen as the starting point for pan angle evaluation (Fig 7). In fact, given a point $p(x, y)$ and the catadioptric image center $C(x_0, y_0)$, we can compute the pan angle θ_p in the omnidirectional referential:

$$\theta_p = \arctan \left(\frac{y - y_0}{x - x_0} \right) \quad (2)$$

$$\tan(\alpha_{pan}) = \frac{v_y}{v_x} \quad (3)$$

The detailed explanation and formulae is inspired from paper [7].

4.2 Face Detection with PTZ Camera

As explained in the previous sections, the catadioptric sensor is able to detect and track face ROIs. By using the ROI data localization from the catadioptric images, the PTZ camera can detect and also make a zoom in the face ROI. Brief descriptions of the different steps for the face ROI localization are as follow: Firstly, perform Viola and Jones face detection algorithm on the unwrapped catadioptric image to identify and localize the face ROI. Secondly, using the pan angle of the ROI center, calculate the pan angle to be directed to the PTZ camera. We use a constant tilt angle and a minimum zoom value. This enables us to point the PTZ camera on a large area where the face is probably located. Thirdly, perform again, Viola and Jones face detection algorithm to detect the face ROI in the obtained PTZ image. Finally, command the PTZ camera in order to center the face ROI detected in the PTZ image (pan and tilt calculation) and then apply the zoom factor computed according to both image and ROI widths and heights, as expressed below:

$$Zoom = \min \left(\frac{Image_{width}}{ROI_{width}}, \frac{Image_{height}}{ROI_{height}} \right) \quad (4)$$

This last step enables us to obtain a zoomed in face image with high resolution which is useful for face recognition processing.

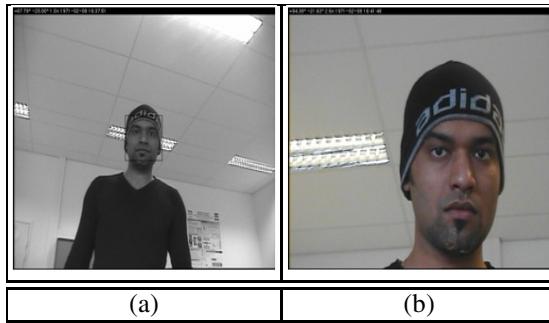


Fig. 7 (a). Detected face Image. (b). Zoomed Image.

5 Face Recognition Based on Fusion of Catadioptric and PTZ Vision Sensors

Face recognition is a very active area of research in computer vision and biometric fields since late 1980s. Among the plethora of techniques available, Eigenfaces technique is one of the earliest appearance-based face recognition methods, which was developed by M. Turk and A. Pentland in [11]. The study and evaluation of the performances of this method for the fusion based system may be interesting to implement in Real Time face recognition.

5.1 *The Eigenface Algorithm*

The eigenfaces technique for face recognition [11], consists of two main phases:

- Learning: This phase uses the idea of the Principal Component Analysis (PCA) and decomposes face images into set of characteristic feature images called eigenfaces.
- Recognition: This phase is then performed by projecting a new face into a low dimensional linear space defined by the generated eigenfaces in order to analyze and then to recognize it.

5.1.1 Learning Phase

The learning phase for face recognition technique involves the following steps:

- The first step is to obtain a set S with M training face images. Let a face image be a two-dimensional N by N array of intensity values. After obtaining our set of face images, we calculate the mean image Ψ .
- The next step consists in calculating the difference faces by subtracting the average face from each input image.

- This set of very large vectors is then subject to the principal component analysis method, which seeks a set of M orthonormal vectors μ_n , which best describes the distribution of the data. The k_{th} vector μ_k is chosen such that : We note that the vectors and scalars are the eigenvectors and eigenvalues, respectively, of the covariance matrix.

$$\lambda_k = \frac{1}{M} \sum_{n=1}^M (\mu_k^T \phi_n)^2 \quad (5)$$

$$\mu_l^T \mu_k = \begin{cases} 1, & l = k \\ 0, & otherwise \end{cases}$$

where Φ_n represents the difference faces by subtracting the average face from each input image : $\phi_n = \Gamma_n - \Psi$ and Ψ represents the mean image :

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n$$

The covariance matrix C is defined as follows:

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T \quad (6)$$

$$\text{With } A = [\Phi_1 \Phi_2 \Phi_3 \dots \Phi_M]$$

Following this equation set, we construct the M by M matrix $L = A^T A$, where $L_{mn} = \Phi_m^T \Phi_n$ and find the M eigenvectors v_n of L . These vectors determine linear combinations of the M training set face images to form the eigenfaces μ_n :

$$\mu_n = \sum_{k=1}^M v_{nk} \Phi_k = Av_n, n = 1 \dots M \quad (7)$$

With this method, the calculations are greatly reduced from the number of pixels in the images (N^2) to the other number of images in the training set (M).



Fig. 8 Average Images

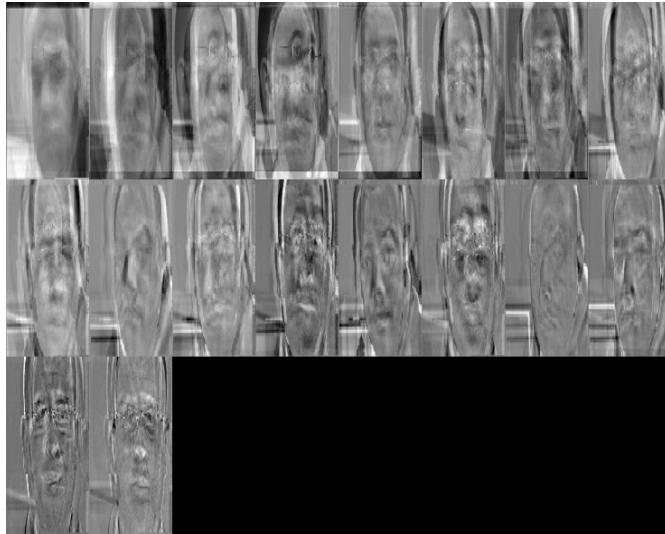


Fig. 9 Output of Eigenfaces

5.1.2 Recognition Phase

The usefulness of the eigenvectors varies according to the associated eigenvalues. In fact, we choose only the most meaningful eigenvectors corresponding to the highest eigenvalues and we ignore the rest. As a consequence, the number of basis functions is reduced from M to M' (M' less than M) and the computation is reduced. The recognition procedure is summarized as follows [11] :

1. A new face, to be identified, is transformed to its eigenface components. Firstly, we compare our input image with the mean image and then multiply their difference with each eigenvector of the obtained L matrix. Each value would represent a weight and would be saved on a vector Ω :

$$\omega_k = \mu_k (\Gamma - \Phi) \quad (8)$$

where $\Omega^T = [\omega_1, \omega_2, \omega_3, \dots, \omega_M]$

2. We determine which face class provides the best description for the input image. This is done by minimizing the Euclidean distance and also the distance between the image and the face space :

$$\xi_k = \|\Omega - \Omega K\|^2 \quad \xi = \|\Phi - \Phi_f\|^2 \quad (9)$$

3. We choose a threshold θ_e that defines the maximum allowable distance from any face class, and a threshold θ_ϵ that defines the maximum allowable distance from face space. For each new image to be identified, calculate its pattern vector Ω , the distance ε_i to each known class, and the distance ε to face space. If the

minimum distance $\varepsilon_k < \theta_e$ and the distance $\varepsilon < \theta_e$, classify the input face as the individual associated with class vector Ω_k . If the minimum distance $\varepsilon_k > \theta_e$ but distance $\varepsilon < \theta_e$, then the image may be classified as "unknown".

4. If the new face is classified as known individual, this image may be added to the original set of familiar set images, and the eigenfaces may be recalculated. This gives the opportunity to modify the face space as the system encounters more instances of known faces.

The entire eigenface program is implemented in C language by using OpenCV library and the main functions developed to learn and recognize faces with OpenCV eigenface methods.

5.2 NOBA Database

In NOBA Face DataBase, we have 10 subjects (or persons). The face data are captured from 2 sensors : Pan Tilt Zoom (PTZ) camera & Camera with infrared light (2D digitizer). During building the Face database, we take into account variations of Poses of head and eyes direction (various poses), Facial expressions (various expressions), Illuminations conditions (various illuminations). The combination of expressions under illumination and poses under expressions, 4 classes are defined :

- C1 : Facial expressions descriptors,
- C2 : Mouth movements descriptors,
- C3 : Pose of head and eye direction descriptors,
- C4 : Various Illuminations.

16 face images for each subject with 4 frontal images, 8 images for different tilts, 2 facial expressions images, 2 images with various illuminations. Image features are mentioned in Table 1. Our idea is to test and verify the results of Eigenface on our own database and see the efficiency and accuracy in terms of implementation.

5.3 Results Obtained under the Heterogeneous Database

To test performances of the developed application, we used a free publicly available face database, the Olivetti Research Laboratories (ORL). This face database provides 10 sample images of each of 40 subjects. For some of the subjects, the images were taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All the images are taken in front of a dark homogeneous background and the subjects are in up-right, frontal position (with tolerance for some side movement). And we merged our subjects (under the same conditions) with this database to obtain a Heterogeneous database. The experimental results are performed with 12 subjects (10 ORL subjects + 2 NOBA database subjects) for the time being. Better results and testing have to be done on extended number of subjects in the near future.

Table 1 Noba Data Base

Image Features
Image 1: Emotional Expression of Smiling
Image 2: Laughter
Image 3: Anger
Image 4: Surprise
Image 5: Eyes Closed
Image 6: Fcae with high illumination
Image 7: Tilt UP 30°
Image 8: Tilt Right 30°
Image 9: Tilt Left 90°
Image 10: Tilt Left 60°
Image 11: Tilt Left 30°
Image 12: Tilt Right 60°
Image 13: Tilt Right 90°
Image 14: Tilt Down 30°
Image 15: Frontal View Neutral Face
Image 16: Face with Low illumination

Various Scenarios

Results obtained show that the developed eigenface program recognizes correctly the considered people belonging already to the face database (training data). The index of the nearest face image corresponds to the ground truth and the confidence level is at 100%. The confidence level is calculated based on the Euclidean distance, so that similar images should give a confidence between 0.5 to 1.0, and very different images should give a confidence between 0.0 to 0.5. The confidence level has to be higher than 0.6 because face images of this person are included to the database but they are not acquired in the same conditions. Excluded images outside of the training database with a confidence level varying from 0.87 until 0.95. Sometimes, this level confidence can be higher until 0.99 and lower until 0.6.

Although the testing images don't correspond to any person, the algorithm provides a confidence level very high (greater than 0.8) in both cases. With results obtained in this test and several tests performed, we conclude that the confidence level is not a good criterion on which we have to rely, in order to recognize people. The algorithm works only when the index of the nearest training face image corresponds exactly to the one of the testing one. The results for the test images (which are mentioned in Table 2 and 3) of two subjects which are trained but does not have the exact match. So confidence level is not 1.

Table 2 Results of test subjects s45 and s50 with non-trained faces

Nearest	Truth	Wrong/Right	Confidence
50	50	correct	0.925854
45	45	correct	0.938389
50	50	correct	0.882469
50	50	correct	0.945845

Training Subjects s20 to s30 and two of our s45 & s50 are trained except s23/6, s23/7, s50/8, s50/9, s45/2, s50/2 & s26/1. The results obtained shows the difference in confidence level depending on whether that face was present in the training data or not. Better the match, higher the confidence level approaching 1.

Table 3 Results of different trained and non-trained faces

Nearest	Truth	Wrong/orrect	Confidence
50	50	Correct	0.924261
45	45	Correct	1.000000
23	23	Correct	0.971391
26	23	Correct	0.936796

Accuracy can be seen to be almost closing 100% as the trained data is able to recognize the test image if the subject is trained previously, and this, no matter if the image do not match exactly.

6 Results

The preliminary results of the actual implementation of the initial prototype of the system is discussed in this section. In order to test the proposed system, the omnidirectional camera has been placed at 20 cm under the dome Axis PTZ camera [7]. Viola Jones [12] algorithm can detect only frontal and semi-profile faces. Once the face ROI localization is determined from the unwrapped image, the program sends the corresponding pan angle via http sockets. Then, a separated process is created to detect the face in the current PTZ image (Fig.7a) by applying Viola and Jones algorithm. Figure 7(a) illustrates the resulting image after performing this face detection algorithm. Finally, the program controls automatically the camera in order to center and zoom in the detected face localization. As a consequence, we obtain a high resolution face picture (Fig.7 b). The localized PTZ image is then downsampled and converted to grayscale image of 92x112 (low-dimensional linear subspace defined by eigenfaces) and saved into the test database for recognition and eigenface

algorithm is implemented to check for the authenticity or match of the given subject with that of the training or registered database or subject. A new face is compared to known face classes by computing the distance between their projections onto the face subspace.

Test results obtained demonstrates that the algorithm provides good results when the testing images correspond to known people belonging to the training face database. Further one system limitation is represented by the poor resolution of the unwrapped images. This can affect performances of the face detection algorithm. Actually, at relatively high distances from the optical axis of the omnidirectional camera (more than 2 meters), Viola and Jones detector is not able to detect face location in these unwrapped images. We can notice that the Viola and Jones algorithm performs more efficiently when the target face is closer to the catadioptric sensor and also when assuming adequate illumination conditions. The current system is also limited by the large number of threads, the program runs. This affects the processing time of the application and the frame rate of the video stream. Therefore, to further optimize these parameters, improvements in the system architecture and improvements related to the decision criteria should be implemented during the last phase of recognition are objectives of fast upcoming studies.

7 Conclusion

In this paper, we proposed a unique vision system, efficient to automatically detect and track ROI at a higher zoom level. Experimental results using robust calibration methods and real-time detection and tracking algorithms demonstrates a significantly improved accuracy in providing a closer look of the target for recognition purposes. A panoramic FOV eliminates the need for more cameras or mechanically turnable camera. The integration of authentication processes like face detection, tracking and recognition makes the system self sustained for biometric authentication. The advantages of omnidirectional sensing are obvious for application like surveillance and immersive telepresence. Our future work focus on the improvement of the current system architecture in order to make the detection and tracking algorithms more robust and faster. Better and efficient methods for object detection in the unwrapped pictures and recognition of face with higher and faster recognition rate are the heart of forthcoming studies. Moreover, our upcoming work will be oriented toward merging more biometric analysis like iris and gait to have a robust and mobile multimodal biometric system.

Acknowledgements. Acknowledgement is made to EU and Upper Normandy region (FRANCE) for the support of this research through the European Interreg IVA project No 4051. The authors would like to thank all people who are given volunteers for the construction of NOBA Database. We thank them for their availability during image data acquisition.

References

1. Baker, S., Nayar, S.K.: A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision* 103(3), 175–196 (2006)
2. Barreto, J.: A unifying geometric representation for central projection systems. *Computer Vision and Image Understanding* 103(3), 208–217 (2006)
3. Boutteau, R.: Reconstruction tridimensionnelle de l'environnement d'un robot mobile à partir d'informations de vision omnidirectionnelle pour la préparation. PhD Thesis., University de rouen (2009)
4. Scaramuzza, R.S.D., Martinelli, A.: A flexible technique for accurate omnidirectional camera calibration and structure from motion. In: Proceedings of the International Conference on Computer Vision, pp. 45–52 (January 2006)
5. Freund, Y., Schapire, E.: A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Science* (1997)
6. Geyer, C., Daniilidis, K.: A unifying theory for central panoramic systems and practical implications. In: Vernon, D. (ed.) *ECCV 2000. LNCS*, vol. 1843, pp. 445–461. Springer, Heidelberg (2000)
7. Iraqui, A., Dupuis, H.Y., Boutteau, R., Ertaud, J.-Y., Savatier, X.: Fusion of omnidirectional and ptz cameras for face detection and tracking. In: International Conference on Emerging Security Technologies, EST (2010)
8. Levenberg, A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics* 2, 164–168 (1994)
9. Mei, C., Rives, P.: Single view point omnidirectional camera calibration from planar grids. In: Proceedings of the International Conference on Robotics and Automation (ICRA), pp. 3945–3950 (2007)
10. Ramalingam, S.R., Lodha Surn, P.: Towards complete generic camera calibration. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp. 767–769 (June 2005)
11. Turk, M., Pentland, A.: Eigenfaces for recognition. In: Vision and Modeling Group The Media Laboratory (1994)

Eigenfaces, Fisherfaces, Laplacianfaces, Marginfaces – How to Face the Face Verification Task

Maciej Smiatacz

Abstract. This paper describes the exhaustive tests of four known methods of linear transformations (Eigenfaces, Fisherfaces, Laplacianfaces and Marginfaces) in the context of face verification task. Additionally, we introduce a new variant of the transformation (Laplacianface + LDA), and the specific interval-based decision rule. Both of them improve the performance of face verification, in general, however, our experiments show that the linear transformations are of marginal importance in this field.

1 Introduction

Automatic face recognition has been a hot research topic for the last 25 years. During this period countless scientific publications appeared on the subject and the early, often quite naive, concepts have been replaced by highly efficient and sophisticated commercial systems. Although there is still a room for improvements in this area (in terms of speed, reliability and security for example), face recognition companies are not very much interested in sharing their experiences and, on the other hand, the scientists, driven by the need to be original rather than useful, have thrown so many ideas into this field, and applied so many, sometimes exotic, mathematical mechanisms, that it is probably going to take years until the well-established knowledge about face recognition technology is available.

In this work we take a closer look at selected *appearance-based* methods of face recognition. Their common denominator is that they operate on the original face images, treating individual pixel intensities as features. They do not rely on any texture or shape descriptors, but use the image data directly. In the typical case,

Maciej Smiatacz

Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland
e-mail: slowhand@eti.pg.gda.pl

the first step of the algorithm concatenates the rows (or columns) of the source bitmap, thus converting the n by m array of pixels into the feature vector containing $N = n \times m$ elements. Then the goal is to find the transformation (usually linear) from this extremely high-dimensional space to the new one, in which the number of coordinates is much smaller, and the important information is not only preserved, but also represented in such a way that the efficient classification is possible. The new space is most often spanned by the eigenvectors of some matrix; the way the matrix is constructed is what distinguishes one method from the others.

Everything began with the famous Eigenface approach [8] from 1990, which was simply based on Principal Component Analysis. At the time of publication the results that it provided were very promising, and the method itself is so well-known and straightforward, that it serves as a reference solution even today. The original publication indicated that the eigenvectors can be visualized as two-dimensional images showing ghost-like faces. This allowed researchers to “look inside” the PCA method. Additionally, the paper contained the algorithm that was able to construct the eigenvectors of $N \times N$ matrix from the eigenvectors of much smaller matrix with dimensions $M \times M$ (where M is the number of training images).

PCA creates a feature space in which the representation of samples is as compact as possible - the higher dimensions can be dropped at the minimal cost (with the minimal loss of information). It does not, however, take the class labels into account. Therefore, from the classification point of view, it is more favorable to employ the Linear Discriminant Analysis (LDA), which uses the between-class scatter matrix as well as the within-class scatter matrices. In the context of appearance-based method, however, the necessary matrix inversion is practically unfeasible. As a solution to this problem the authors of the Fisherface method [1] proposed to reduce the dimensionality of the samples by using PCA before applying the discriminant analysis. Interestingly, it was hardly an original concept - the same operation was mentioned by Sammon in his paper [4] from 1970. Nevertheless, Fisherfaces gained a lot of interest and later on PCA became a standard preprocessing algorithm for many other methods.

While the goal of the Fisherface approach is to create the space that emphasizes the differences between classes, the Laplacianface method [2] tries to learn a locality preserving face subspace which is insensitive to outlier and noise. Finally, the criterion used in Marginface algorithm [10] favors the solutions that keep the patterns from the same class as close as possible, while pushing the patterns from other classes away - therefore the margin between classes is maximized.

Although many works on face recognition discuss the problem of face *classification*, in practice we very rarely construct biometric systems that classify persons. If the task of an application is to grant a person the access to some resource, the application needs to *verify* the identity of the person. In other words, we do not want the algorithm to pick one of the known class labels. The class label (e.g. the name of the person) is given, and the algorithm has to check whether the presented biometric pattern is *similar enough* to the stored template. This is a serious change of perspective. At first it might seem that we need the data describing one class only - after all, when we want to train a biometric system on a new laptop, for example,

we enroll only our own images or fingerprints. However, the abovementioned space transformation methods are generally useless in the case like that - they also need some information about “the rest of the world” to distinguish between us and other people. Thus, in addition to the positive set (the images showing the face of the legitimate user) we also need the set of negative examples (pictures of other persons). Still, it is not clear how to construct the negative set, i.e. how big it should be or what images should be used. After that two scenarios are possible: we can either treat the verification task as a binary classification problem, or we can use the negative set only to define the transformation, and then ignore the information about the other class, applying some arbitrarily set similarity threshold.

While working on our SART-2 project [5] we tested, among others, the Eigenface, Fisherface, Laplacianface and Marginface methods using our CF framework [7]. The evaluation was fully unbiased as we did not want to promote any new approach; our goal was just to check the usefulness of the methods for the face verification task. In our opinion such a comparison may be of interest, especially for practitioners working in the field of biometrics.

2 Compared Methods

In this section the four appearance-based methods are briefly presented. In each case we assume that the training set contains M column vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$; M_P of them represent the positive training images (P class), and M_N describe the negative samples (N class). Each method creates some transformation matrix \mathbf{U} that is employed to convert the \mathbf{x}_i training vectors into the output vectors \mathbf{y}_i , which in turn can be passed to the decision-making module.

2.1 Eigenfaces

Every input vector \mathbf{x}_i of the Eigenface method [8] is constructed directly from the original image by concatenating the subsequent columns of the training bitmap. As a consequence, at the beginning of the process source images are usually scaled down to the size of about 50×50 pixels ($N = 2500$), to prevent the \mathbf{x}_i vectors from being too large. Having calculated the global mean of all training samples, μ_T , we are able to create the following covariance matrix:

$$\mathbf{C} = \frac{1}{M} \sum_{i=1}^M (\mathbf{x}_i - \mu_T)(\mathbf{x}_i - \mu_T)^T = \frac{1}{M} \sum_{i=1}^M \phi_i \phi_i^T = \mathbf{A} \mathbf{A}^T \quad (1)$$

where $\phi_i = \mathbf{x}_i - \mu_T$, and $\mathbf{A} = [\phi_1 \ \phi_2 \ \dots \ \phi_M]$. The transformation matrix \mathbf{U}_E is composed from the K first eigenvectors \mathbf{u}_j of matrix \mathbf{C} , i.e. $\mathbf{U}_E = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_K]$. The dimensions of \mathbf{C} , however, are usually large ($N \times N$), and the maximal number of meaningful eigenvectors (with associated eigenvalues greater than 0) is equal to M ,

where, in most of the cases, $M \ll N$. Thus, it is much more effective to compute the eigenvectors \mathbf{v}_j of the smaller $M \times M$ matrix $\mathbf{A}^T \mathbf{A}$, and then calculate the eigenvectors \mathbf{u}_j of matrix \mathbf{C} by using the formula

$$\mathbf{u}_j = \sum_{s=1}^M v_{js} (\mathbf{x}_j - \mu_T) \quad (2)$$

where v_{js} denotes the s -th element of vector \mathbf{v}_j .

The dimensionality of the new feature space, K , is rarely set to the maximal possible value. Instead, K is calculated on the basis of eigenvalues of \mathbf{C} so that the transformation preserves certain amount (e.g. 98%) of the most important information. Finally, we obtain the output vectors:

$$\mathbf{y}_i = \mathbf{U}_E^T (\mathbf{x}_i - \mu_T) \quad (3)$$

As a result, the representation of the data is much more compact and \mathbf{y}_i vectors have zero mean. It is important to remember, however, that the axes of the new space correspond to the directions of *maximal variance* observed within the training set.

2.2 Fisherfaces

Although it is theoretically possible to apply the Linear Discriminant Analysis (based on the Fisher criterion) to the original, N -element input vectors, in practice the within-class scatter matrix, which must be inverted, is always singular in such a case (due to the small sample size problem). Therefore, the dimensionality of the vectors is reduced by using PCA first; in other words, the output vectors of Eigenface method are treated as input vectors \mathbf{x}_i of the LDA. This is the main concept behind the so-called Fisherface method [1].

First we calculate the mean μ_l of each of the L classes (for the verification task $L = 2$), and then the between-class scatter matrix

$$\mathbf{B} = \sum_{l=1}^L M_l (\mu_l - \mu_T) (\mu_l - \mu_T)^T \quad (4)$$

and the within-class scatter matrix

$$\boldsymbol{\Sigma} = \sum_{l=1}^L \sum_{\mathbf{x}_i \in C_l} (\mathbf{x}_i - \mu_l) (\mathbf{x}_i - \mu_l)^T \quad (5)$$

where C_l denotes the l -th class and M_l is the number of images in this class. If PCA was applied as a preprocessing method, then $\mu_T = 0$.

The columns of the transformation matrix \mathbf{U}_F are the eigenvectors \mathbf{u}_j of the matrix $\boldsymbol{\Sigma}^{-1} \mathbf{B}$, corresponding to its K largest eigenvalues. The transformation is then very simple (the input samples have zero mean already):

$$\mathbf{y}_i = \mathbf{U}_{\mathbf{F}}^T \mathbf{x}_i \quad (6)$$

The resultant feature space maximizes the ratio of between-class scatter to the averaged within-class scatter. We have to bear in mind, however, that this is a linear transformation that implicitly treats all classes as having the same covariance matrix Σ . Moreover, the overall concept makes little sense for the classes with non-Gaussian distributions. Additionally, there are at most $L - 1$ nonzero eigenvalues of $\Sigma^{-1}\mathbf{B}$, thus the verification (with $L = 2$) becomes a one-dimensional problem. This means an enormous information reduction that may significantly affect the method's performance.

2.3 Laplacianfaces

This method also uses the output of the Eigenface algorithm as the source dataset. The goal of the Laplacianface approach [2] is to find the transformation that preserves the local information - in contrast to PCA that protects the overall structure of the image space. First of all, the global $M \times M$ similarity matrix \mathbf{S} is created. If \mathbf{x}_i is among the k nearest neighbors of \mathbf{x}_j , then the S_{ij} and S_{ji} entries of the matrix are defined as follows:

$$S_{ij} = S_{ji} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|/t) \quad (7)$$

where t determines the rate of decay of the similarity function and $\|\cdot\|$ denotes the L^2 norm; otherwise $S_{ij} = 0$. Then we construct the diagonal matrix \mathbf{D} whose entries are the row sums of \mathbf{S} , i.e. $D_{ii} = \sum_j S_{ji}$. The matrix $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is the so-called Laplacian matrix. If we treat each input vector \mathbf{x}_i as a column of the $N \times M$ matrix \mathbf{X} then we can create the following matrices

$$\Lambda = \mathbf{XLX}^T \quad \Delta = \mathbf{XDX}^T \quad (8)$$

and then use the eigenvectors of $\Lambda \Delta^{-1}$ as columns of the transformation matrix \mathbf{UL} . The output vectors \mathbf{y}_i are calculated with the formula analogous to (6). It is worth noticing that the method ignores the class labels of the training samples, thus in the case of multi-modal distributions there is a danger that the transformation will emphasize the similarities between the corresponding modes of different classes (representing, for example the same pose or illumination conditions), rather than the similarities of patterns coming from the same class.

2.4 Marginfaces

Marginface [10] is a method that uses the average neighborhood margin maximization (ANMM) and exists in two versions: the linear (considered here) and the kernelized one. The authors of the original paper do not mention any preprocessing

with PCA, they simply resize the input images to the size of 32×32 pixels and then convert them to the \mathbf{x}_i vectors. For each of them we have to find the “homogeneous neighborhood” N_i^o (the set of the p nearest neighbors from the same class) and the “heterogeneous neighborhood” N_i^e (the set of the r nearest neighbors from other classes). The similarity of the neighbors is measured with the help of L^2 norm. Next we construct the “scatterness matrix”

$$\mathbf{S} = \sum_{\substack{i,k: \\ x_k \in N_i^e}} \frac{1}{r} (\mathbf{x}_i - \mathbf{x}_k)(\mathbf{x}_i - \mathbf{x}_k)^T \quad (9)$$

and the “compactness matrix”

$$\mathbf{C} = \sum_{\substack{i,j: \\ x_j \in N_i^o}} \frac{1}{p} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \quad (10)$$

The columns of the transformation matrix \mathbf{U}_M are the eigenvectors of $\mathbf{A} = \mathbf{S} - \mathbf{C}$. However, only those of them which correspond to the non-negative eigenvalues are taken into account. The operation similar to (6) is applied to obtain the output \mathbf{y}_i vectors.

3 Decision Making

The appearance-based methods described in the previous section use global image information (pixel intensities) to construct a new feature space, trying to extract some intrinsic knowledge from the limited number of training samples. They do not, however, include any decision rules. The goal of this work is not to create the fully functional biometric system, but to compare the four methods of linear transformations. Therefore, we resort to the simplest decision-making algorithms, and the use of the Euclidean distance to the class mean as the similarity measure seems to be the first natural choice. In the case of the verification we have two options, related to the two questions we may ask: 1) is the test image \mathbf{x} similar enough to the P class or not, 2) is the test image more similar to the class P than to the class N. The first question translates into the following decision rule:

$$\mathbf{x} \in P \quad \text{if} \quad \|\mathbf{x} - \mu_P\| < \alpha \cdot w \quad (11)$$

where μ_P is the mean of the positive class, α is the maximal distance from μ_P , observed in the training set of P class, and w is a parameter. This operation reduces to the simple thresholding. Although we do not use the information about the negative class in the decision rule, we have to remember that this knowledge might have been used for the construction of the linear transformation. We should note, however, that the decision surface defined by (11) is a hypersphere, i.e. we have the same threshold α for all directions, while some of the methods, e.g. PCA, will try to

maximally differentiate the variances along the axes of the new space. Obviously, the use of Mahalanobis metrics might help, but this would lead us once again to the small sample size problem in P class. Therefore, to make use of the information about the within-class scatter of P without the need to invert the covariance matrix, the following interval-based decision rule is proposed:

$$\mathbf{x} \in P \text{ if } \forall s \quad -\alpha_s \cdot w + x_{s\min} < x_s < x_{s\max} + \alpha_s \cdot w \quad s = 1..K \quad (12)$$

where x_s is the s -th element of input vector \mathbf{x} (s -th feature value), $x_{s\min}$ and $x_{s\max}$ are the minimum and maximum values of s -th feature observed in the training set, $\alpha_s = |x_{s\min} - x_{s\max}|$, and w is a parameter.

Asking the second question leads to the binary classification problem:

$$\mathbf{x} \in P \text{ if } \|\mathbf{x} - \mu_P\| < \|\mathbf{x} - \mu_N\| \quad (13)$$

where μ_N is the mean of the negative class. In practice, however, the decision rule (13) is useless, as the within-class scatter of N is usually huge in comparison with P, which results in extremely high false acceptance rates.

4 Experiments

The experiments were performed with the help of our CF framework [7] that facilitates the training and testing of complex pattern recognition systems. It is a modular system, into which the independently developed C++ units (such as normalizers, transformers or classifiers) can be plugged easily. One of the advantages of CF is that it supports batch operations (e.g. creating and evaluating the verification systems for subsequent classes) on large datasets.

It is a well-known fact that appearance change of face images due to pose and illumination changes is usually larger than that caused by different identities. Different poses and lighting directions make class distributions multimodal; the corresponding modes from different classes are situated closer to each other than different modes of the same class. In our opinion (despite the numerous experiments described in the literature) application of linear transformations in such a case makes no sense at all. Thus, in our tests we used approximately frontal images captured under constant illumination conditions. The photographs came from the CMU-PIE database [6]. For each of the 68 persons the images taken from two almost frontal viewpoints (cameras no 7 and 27) during the expression session (showing the subjects with a neutral expression, smiling, and blinking) were selected as the P set. The number of positive training images per class ranged from 4 to 8. The negative set was constant for all classes and contained 90 frontal images of different persons, taken from another database (FERET [3]). Finally, the testing set included 4693 nearly frontal images (cameras 7, 9, 27) from CMU-PIE database (around 70 “positive” and 4600 “negative” samples). The training and testing procedure was repeated 68 times for every verification method, i.e. the combination of a linear

transformation and a decision rule. The faces were cropped by the Viola & Jones algorithm [9]. Then the histogram equalization was applied and the input images were scaled down to the resolution of 32×32 for Marginface and 50×50 for other methods.

At first we applied decision rule (11) with $w = 1.1$. The parameters of the Laplacianface were set to $k = 4$, $t = 100$. We also introduced a new transformation comprised of Laplacianface followed by LDA. Three settings of Marginface were tested, as shown in Table 1 that contains the results of the first experiment.

Table 1 Average and extreme false rejection and false acceptance rates, achieved after the whole test (68 classes \times 4693 images) for different verification methods that used decision rule (11); p and r values are given for Marginface method

	Eigen	Fisher	Laplacian	Laplacian	Margin	Margin	Margin	Euclidean
			+ LDA	(1,1)	(4,5)	(4,10)		
FAR (%)	8.03	1.28	8.03	1.17	4.79	11.70	12.25	0.49
FRR (%)	14.22	30.30	14.22	28.53	20.87	12.19	14.34	19.32
max FAR	46.07	11.24	46.07	13.05	58.59	69.02	78.83	8.52
min FAR	0	0	0	0	0	0	0	0
max FRR	91.43	92.54	91.43	92.54	92.42	91.30	92.42	92.42
min FRR	0	0	0	1.52	0	0	0	0

As we can see, Marginface is ineffective, Laplacianface and Eigenface perform identically but their false acceptance rates are much too high, and our combination of Laplacianface and LDA provides the lowest average FAR with acceptable FRR. However, if we take look at the last column, showing the results of the most naive algorithm (no linear transformation, just pixel by pixel comparison with the mean), we realize that all the methods are uniformly useless. This could be partially caused by the fact that the decision rule (11) is incompatible with some of the transformations. Thus, we repeated our experiments using (12) with $w = 0.5$. Table 2 shows the results.

Indeed, if we treat FAR as a more important parameter than FRR, then the introduction of (12) is a significant improvement, particularly in the case of the Laplacianfaces (for most of the classes FAR = 0). The FRR became much higher, however, so it is not clear, whether this method is truly better than the most naive solution based entirely on (11).

Analyzing the average FAR and FRR coefficients one could think that there is some potential at last in some of the linear transformations. Unfortunately, if we take a look at the extreme values, it is obvious that despite the class-specific adjustments of decision rules (expressed by α or α_s parameters), the performance hugely varies from person to person, sometimes reaching totally unacceptable levels (e.g. $FRR \approx 93\%$ - remember that pose and illumination variations are minimal).

Table 2 Average and extreme false rejection and false acceptance rates, achieved after the whole test (68 classes \times 4693 images) for different verification methods that used decision rule (12)

	Eigen	Laplacian	Margin	Euclidean
	$(p = 4, r = 5)$			
FAR (%)	0.29	0.04	0.66	0
FRR (%)	48.96	55.42	48.06	90.48
max FAR	10.27	0.76	10.81	0
min FAR	0	0	0	0
max FRR	93.85	92.42	93.85	93.85
min FRR	1.52	4.35	2.94	88.24

5 Conclusions

We independently tested four well-known methods of linear transformations in the context of the face verification task. We also proposed a new decision rule and the new variant of the transformation that combines the Laplacianface method with the Linear Discriminant Analysis. The tests were exhaustive and we tried to avoid non-sense challenges, like the evaluation of the linear transformation performance in the presence of pose and illumination changes. Unfortunately, we cannot confirm that the linear transformations are useful when it comes to appearance-based face verification. In fact their importance is marginal, if any. The practitioners willing to create the real-life face verification software should concentrate on much more significant factors, such as the precise face localization and cropping, illumination normalization, pose estimation, or invariant local texture features. Our experiments show that the “magical” power of the eigenvectors of yet another matrix is not able to solve any practical problem in the face verification field.

References

1. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19(7), 711–720 (1997)
2. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.-J.: Face Recognition Using Laplacianfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence* 27(3), 328–340 (2005)
3. Phillips, P.J., Moon, H., Rauss, P.J., Rizvi, S.: The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(10), 1090–1104 (2000)
4. Sammon, J.W.: An Optimal Discriminant Plane. *IEEE Trans. Computers* C-19(9), 826–829 (1970)
5. SART-2 project (2012), <http://sart2.eti.pg.gda.pl/>

6. Sim, T., Baker, S., Bsat, M.: The CMU Pose, Illumination, and Expression (PIE) Database. In: Proc. 5th Int. Conf. on Automatic Face and Gesture Recognition (2002)
7. Smiatacz, M., Przybycien, K.: A Framework for Training and Testing of Complex Pattern Recognition Systems. In: IEEE Conf. on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, pp. 198–203 (2011)
8. Turk, M., Pentland, A.: Eigenfaces for Recognition. *J. Cognitive Neuroscience* 3(1), 71–86 (1990)
9. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. *Int. J. Comp. Vision* 57(2), 137–154 (2004)
10. Wang, F., Wang, X., Zhang, D., Zhang, C., Li, T.: marginFace: A novel face recognition method by average neighborhood margin maximization. *Pattern Recognition* 42, 2863–2875 (2009)

A Content Based Feature Combination Method for Face Recognition

Madeena Sultana and Marina Gavrilova

Abstract. In the last few years, Content Based Image Retrieval (CBIR) system, where images are searched based on their visual contents instead of annotated texts, has drawn enormous attention of researchers because of its growing demand from real world applications. According to many, face recognition is one of the most potential applications of CBIR. In this paper, a content based face recognition process, where color, texture, and shape features are combined to enhance the retrieval accuracy of the system, is proposed.

Keywords: Face Recognition, Content Based Image Retrieval (CBIR), Color Histogram, Gabor Filter, Histogram Intersection.

1 Background and Motivation

With the advancement and the availability of image capturing and storage devices, the collection of digital images is growing rapidly, day by day. To meet the increasing demand of searching, browsing, and recognizing images from the large collection of databases, a number of image retrieval systems have been developed over the past two decades. The retrieval systems can be broadly classified into two categories [16]: text based [2] and content based [9]. In text based approach, textual features such as filenames, captions, and keywords are used to annotate and retrieve images. However, researchers explored a number of disadvantages relating to text based query such as human intervention, manual annotation, domain expertise, language barrier etc. To overcome the aforementioned disadvantages of traditional text based retrieval system, Content Based Image Retrieval (CBIR) [9] system was introduced in 1992. In CBIR, images are searched based on the visual contents e.g. color,

Madeena Sultana · Marina Gavrilova

Dept. of Computer Science, University of Calgary, Calgary, AB, Canada

e-mail: {msdeena, mgavrilo}@ucalgary.ca

texture, and shapes. Therefore, it produces more relevant result than text based image retrieval systems. Moreover, CBIR is also efficient in terms of computation time since color, shape, and textures are low level image features and can be extracted by using simple and fast algorithms.

At present, biometric systems such as, face recognition [12], iris recognition [21], fingerprint recognition [20], body motion recognition [14] have also an increasing demand for various applications ranging from entertainment to law enforcement. Among all, face detection and recognition from databases has received significant attention from both the academic and industrial communities because of its availability and non-contact acquisition process. Nevertheless, an efficient and robust face recognition system is still an unsolved problem. Many researchers mentioned that CBIR can be an efficient technique for face recognition. However, very few papers have been published on CBIR based face recognition. This instigated us to explore the challenges/obstacles of CBIR based face recognition system and to develop an efficient content based system to recognize face images.

2 Literurture Review

Because of the escalating demand of browsing and recognising digital images in different applications, CBIR remains as a hot topic of research since its origins. As a result, a good number of commercial products for image retrieval from databases based on CBIR are available on the market. QBIC [4], Virage [6], SIMPLICITY [19] are some examples of such systems. However, content based systems are comparatively new approach and therefore, are under investigation. We classified the trend of current research works on CBIR into two major categories: 1. general purpose CBIR and 2. specialized or domain specific CBIR. The latter is the most recent and more challenging area of research.

For general purpose CBIR, researchers are mainly investigating various techniques to improve the performance of the retrieval systems. It is known that performance of the approaches using single features sometimes degrades significantly when large variety of input images is used. Therefore, current research shows a predilection for fusion of different features to boost up the retrieval accuracy. Yue et al. [22] combined color and texture features to enhance the retrieval accuracy of their system. Through experimentation, they proved that integration of color and texture feature improves the recognition performance. Singha and Hemchandra [15] showed that integrated color histogram and wavelet based texture features not only enhances the retrieval accuracy, but also offers ease of computation. However, due to the lack of considering shape features, their method is not invariant to rotation and affine transformation. Based on the above, we integrated color, texture, and shape features to make our system robust, efficient, and invariant to variability in image type, orientation and quality.

Applications intended for general purpose retrieval of images do not produce good results for specific domains. For example, general purpose retrieval system

can distinguish a flower image from a building image, but might be unable to recognize identical twins. Therefore, many researchers are now devoted to develop CBIR systems for very specialized domains [16], where the main challenges are the identification and selection of appropriate feature and feature extraction methods for that specific environment.

Despite the significant demand, content based face recognition system is an understudied problem compared to other applications. Research works mainly done on content based face detection rather than recognition. Very few works are published till date. Nandwal et al. [13] used content based feature matching method, but face features were extracted by traditional Linear Discriminate Analysis (LDA). Another very recent content based approach is proposed by Iqbal et al. [7]. They integrated three content based feature extraction methods by fuzzy heuristic rules. However, none of the aforementioned works are evaluated using any standard face or biometric database.

3 Proposed Method

Face database may contain large varieties of face images having different expression, pose, rotation, scale, facial detail etc. For this reason, use of a single feature is not sufficient to recognize faces accurately. Therefore, in the proposed method, color, texture, and shape attributes are integrated by their weighted combination, in order to improve the performance of recognition system. Also, to reduce computation overhead of multiple features, we have chosen simple and fast extraction techniques of these three features. First of all, because of the simplicity, trivial computation, and robustness against small changes in camera viewpoint and noise, color histogram [17] is chosen to extract color attributes of face images. However, it is known that apparently different images may produce the similar color histogram even if their contents are different. This problem is overcome by fusing image texture and shape features along with color. Secondly, the most widely used texture analysis method Gabor wavelet filter [3] is utilized to extract texture features. Thirdly, we exploited affine moment invariants [5] to make our system invariant to affine transform e.g. any combination of rotation, translation and scaling. In a comparative study, Vadivel et al. [18] proved that histogram intersection performs better than Euclidean distance for color feature matching. So, in our method, the similarity of color features between query and database image is measured by histogram intersection [8]. Another popular and simple method: Euclidean distance, is applied to measure the similarity of texture and shape features. Finally, a ranked list of matched images is retrieved by combining the weighted color, texture, and shape similarity values.

A block diagram of the proposed method is depicted in Fig.1. The shaded blocks indicate the novel components of our proposed method. Each of these components contributes to improve the overall performance of our system. Use of affine moment invariant makes our system tolerant to affine transform and histogram intersection

improves the accuracy of color feature matching. The most significant advantage of weighted combination method is its tuneable weight parameters â€¢ weights can be adjusted according to the performance of each feature regardless of defining any threshold value for matching. Thus, our method can be applied to any database by adjusting the weights only. Moreover, weighted fusion of features ensures fast computation due to its simplicity.

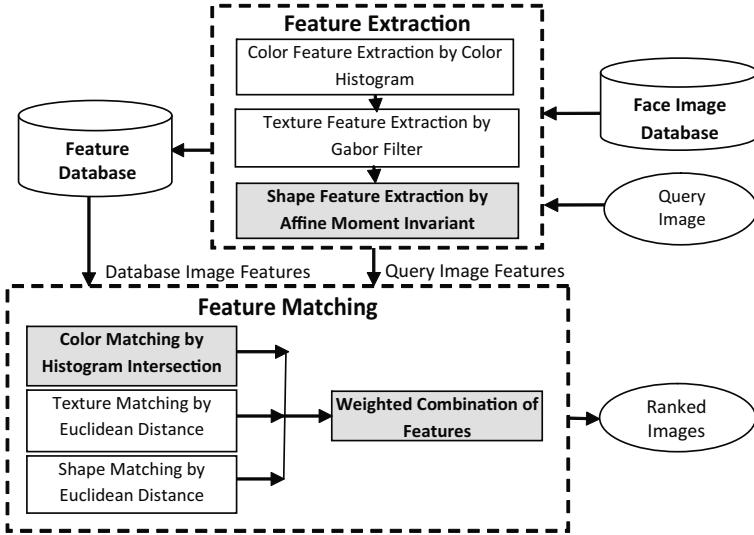


Fig. 1 Block diagram of proposed method

The main two components of the proposed method are Feature Extraction and Feature Matching. Each component is described below:

3.1 Feature Extraction

Color Features: Normalized color histograms $H(i)$ [17] is computed as color feature of query and database images, respectively, as follows:

$$H(i) = \frac{I(i)}{\sum_{i=1}^{256} I(i)}, \quad (1)$$

where $I(i)$ is the intensity value of grayscale image or the value of color components red (R), green (G), or blue (B) of color image. Index i represents a histogram bin or gray level.

Texture Features: We used Gabor wavelet filter to extract texture features from query and database images. The 2D Gabor filter for spatial domain is defined by the following equation [3]:

$$g, \lambda, \theta, \Psi, \sigma, \gamma(x, y) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{y} + \psi\right), \quad (2)$$

where

$$x' = x \cos(\theta) + y \sin(\theta) \quad (3)$$

$$y' = y \cos(\theta) - x \sin(\theta) \quad (4)$$

λ is the wavelength of sinusoidal factor, θ is the orientation of the normal to parallel stripes of a Gabor function, ψ represents the phase offset, σ is the standard deviation of the 2D Gaussian envelope, and γ represents the spatial aspect ratio which specifies the ellipticity of the support of the Gabor function. The database image $T(x, y)$ and query images $Q(x, y)$ are convoluted (*) by the Gabor filter $g(x, y)$ as follows [3]:

$$Qg(x, y) = Q(x, y) * g(x, y) \quad (5)$$

$$Tg(x, y) = T(x, y) * g(x, y), \quad (6)$$

where Q_g and T_g are the outputs of Gabor filter. The energy content can be computed from the magnitude of Q_g and T_g using following equations [3]:

$$E_Q = \sum_x \sum_y |Qg(x, y)| \quad (7)$$

$$E_T = \sum_x \sum_y |Tg(x, y)|, \quad (8)$$

where E_Q is the energy component of query image and E_T is the energy component of database image. The mean and standard deviation of energy components of images at different scale and orientation are considered as texture features in the proposed method.

Shape Features: We propose to exploit the affine moment invariants to achieve invariance of affine transformation of images. Affine transform is linear 2D geometric transformations which maps spatial coordinates (x, y) of an image into new coordinate (u, v) by applying a linear combination of translation, rotation, scaling [5]. Higher order moments are sensitive to noise; therefore, in our proposed method we computed affine moment invariants up to third order. As a result, normalized affine moments are invariant to translation, rotation, and scale changes.

3.2 Feature Matching

We propose to measure the similarity by Jain and Vailayasa ÁŽ histogram matching algorithm. If T_R , T_G , and T_B are the normalized color histograms of database image

T (an image in the database), and Q_R , Q_G , Q_B are the normalized color histograms of the query image Q , then the similarity between Q and T is measured by the following equation [8]:

$$S_H(T, Q) = \frac{\sum \min(T_R(r), Q_R(r)) + \sum \min(T_G(g), Q_G(g)) + \sum \min(T_B(b), Q_B(b))}{\min(|T|, |Q|) \times 3}, \quad (9)$$

where $S_H(T, Q)$ denotes the similarity value of color feature and lies in the interval $[0, 1]$. If the histograms T and Q are identical, then $S_H(T, Q) = 1$. Euclidean distance is applied to calculate the similarity of texture and shape features as follows:

$$S_T(T, Q) = \sqrt{\sum_i (E_T - E_Q)^2} \quad (10)$$

$$S_M(T, Q) = \sqrt{\sum_i (M_T - M_Q)^2} \quad (11)$$

The value of S_T denotes the similarity of texture feature between database image (E_T) and query image (E_Q). Similarly, S_M represents the similarity of shape feature between database image (M_T) and query image (M_Q). The similarity values are then normalized and subtracted from 1. Therefore, if the query image and database image are identical, the value of similarity indices will be equal to 1.

Finally, contributions of the three distinct features are integrated by calculating the weighted sum (S) of S_H , S_T , and S_M as follows:

$$S = \frac{w_h \times S_H + w_t \times S_T + w_m \times S_M}{w_h + w_t + w_m}, \quad (12)$$

where w_h , w_t , w_m are the weights assigned to the similarity indices of the color, texture, and shape, respectively. This novel proposed feature combination technique enhances the probability of the correct image recognition consequently reduces the chance of false recognition.

4 Experimental Result and Discussion

All experiments are carried out on Windows 7 operating system, Intel Core i3 2310M processor with 4GB RAM. Matlab version R2010a is used for implementation and experimentation of the proposed method.

To evaluate the performance of our method we used the following two standard and publicly available datasets:

AT&T Dataset [1]: It contains 400 grayscale images of size 92×112 pixels, with 256 gray levels per pixel, in PGM format. There are 10 different images for each of the 40 distinct subjects. Images were taken at different times, illumination, facial expressions, side movements, and facial details.

AR Dataset [10]: It contains color images of 70 males and 56 females. Each subject has 26 different images in two sessions. Both sessions contain 13 images including different conditions such as natural expression (I_1), smile (I_2), anger (I_3), scream (I_4), different illumination ($I_5 - I_7$), wearing sunglasses in different lightning conditions ($I_8 - I_{10}$), and wearing scarves in varying illumination ($I_{11} - I_{13}$).

From the above two datasets we created following three databases (DB) for our experimentation:

1. **Grayscale database:** Among the 40 subjects of AT&T dataset, 20 subjects were chosen randomly for this database. Therefore, it contains 200 images in total with 10 images per person with variation of pose, expression, rotation, and time.
2. **Color database:** From the cropped images of AR dataset [11], we have chosen 20 subjects randomly from both sessions having varying pose and expressions. Therefore our database contains total 160 images, composed by a subset $D_{color} = \{I_{i1}, I_{i2}, I_{i3}, I_{i4}, I_{i14}, I_{i15}, I_{i16}, I_{i17}\}$ of AR dataset, where $i=1, 2, 3, \dots, 20$.
3. **Critical query database:** This database was created to evaluate our system in critical conditions such as persons wearing sunglasses under different illumination and time. For this purpose, we have created a query database containing total 120 images, $Q_{critical} = \{I_{i8}, I_{i9}, I_{i10}, I_{i21}, I_{i22}, I_{i23}\}$ from AR dataset where $i=1, 2, 3, \dots, 20$. The query images from this database are matched to our color database (D_{color}).

Through experimentation we found that color histogram produced better result in grayscale database and texture feature performed best in color and critical query database. So, in proposed method, the highest weight is applied to color for grayscale database and to texture for color and critical query databases. The weights w_h, w_t, w_m are determined through experimentation. At first, we evaluated the performance of each individual method. Similarly to [16], initial weights are set as $w_h=45, w_t=15, w_m=40$. The weights are then fine tuned to produce the best recognition rate for each individual database. Finally the weights are set to $w_h=47.5, w_t=12.5, w_m=40$ for grayscale and $w_h=40, w_t=12.5, w_m=47.5$ for the two other databases, respectively.

In content based image retrieval system, precision and recall are the standard methods to verify the effectiveness of any technique. The precision and recall are defined as follows:

$$Precision = \frac{I_N}{R} = \frac{TruePositives}{TruePositive + FalsePositives} \quad (13)$$

$$Precision = \frac{I_N}{T} = \frac{TruePositives}{TruePositive + FalseNegatives}, \quad (14)$$

where I_N is the number of images retrieved that are most relevant to query, T is the total number of similar images in the database, and R is the total number of retrieved images. Different researchers use different values for R and T . We considered $T = 10$ for grayscale database and $T = 8$ for color database, since former database contains 10 different images per person and latter database contains 8 images per person, respectively. The 5 topmost images retrieved from database are considered as the search result of any query image, so $R = 5$. Table 1 and Table

2 summarize the recognition performance in terms of average precision and recall value of different methods in different databases.

Table 1 Comparison of average precision of different methods for different databases

Method	Average Precision			
	Grayscale Database	Color Database	Critical Query	Database
Color histogram method [17]	0.95	0.95	0.65	
Affine moment invariant method [5]	0.49	0.72	0.48	
Gabor filter method [3]	0.94	0.97	0.93	
Proposed method	0.98	0.99	0.94	

Table 2 Comparison of average recall of different methods for different databases

Method	Average Recall			
	Grayscale Database	Color Database	Critical Query	Database
Color histogram method [17]	0.48	0.59	0.41	
Affine moment invariant method [5]	0.24	0.45	0.30	
Gabor filter method [3]	0.47	0.60	0.58	
Proposed method	0.49	0.62	0.59	

From Table 1 and Table 2 we can see that the performance of the proposed method by using the weighted combination of color, texture, and shape is better than that of using any of the three methods individually. For critical query database the performance of color histogram significantly degraded because of illumination change and occlusion. However, this does not deteriorate the performance of our proposed method since we assigned higher weights to texture feature for this database. Fig.2 - Fig.4 illustrate some sample retrieval results (ranked) of the proposed method from the different databases.

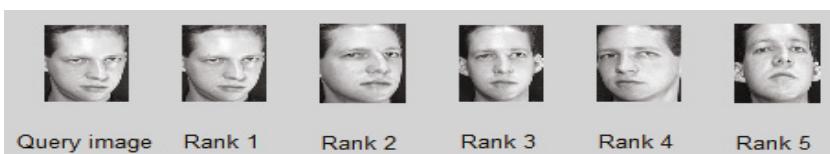


Fig. 2 Person having different pose and rotation is recognized by proposed method from grayscale database

From Fig.2 - Fig.4, we can see that proposed method can recognize person from both grayscale and color images having different pose, expression, sessions, alignment, facial details, illumination variation, and some occlusions.

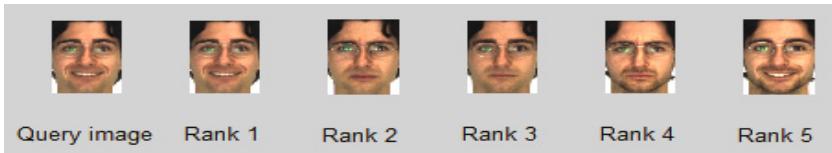


Fig. 3 Person having different expressions at different times is recognized by proposed method from color database

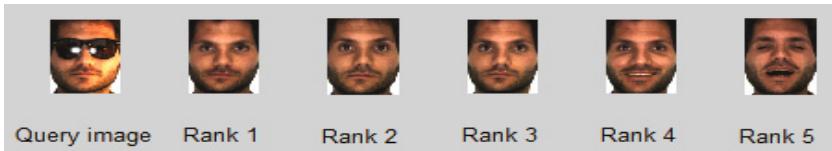


Fig. 4 Critical query image (sunglass, varying illumination and time) is matched in color database by proposed method

5 Conclusion

In this paper, a new content based face recognition method by using integrated color, texture, and shape features has been proposed. Fusion of the three distinct parameters reduces the number of false retrievals as it is highly unlikely that three perceptually different images are assigned high similarity values by corresponding feature. Experimental results show that this method can efficiently recognize face images from standard face databases. Comparative analysis confirms that proposed method has better performance than single feature based methods. Moreover, the performance of our method does not depend on the training dataset since no training session is required to apply this method. We believe our method can be applied to any database effectively because of its high recognition rate, ease of computation, and easy weight adjustment features. Our future works include development of a learning system for auto weight adjustment in different databases, as well as increasing running time investigation by GPU based approaches.

References

1. AT&T Lab. Cambridge (accessed on January 20, 2013),
www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
2. Chang, N.S., Fu, K.S.: A Relational Database System for Images, Tech. Report TREE 79-28, Purdue University (May 1979)

3. Daugman, J.G.: Uncertainty Relations for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-dimensional Visual Cortical Filters. *J. of the Optical Society of America A* 2, 1160–1169 (1985)
4. Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., Equitz, W.: Efficient and Effective Querying by Image Content. *J. Intell. Inf. Syst.* 3(3-4), 231–262 (1994)
5. Flusser, J., Suk, T.: Rotation Moment Invariants for Recognition of Symmetric Objects. *IEEE Trans. Image Proc.* 15, 3784–3790 (2006)
6. Gupta, A., Jain, R.: Visual Information Retrieval. *Commun. ACM* 40(5), 70–79 (1997)
7. Iqbal, K., Odetayo, M.O., James, A.: Content Based Image Retrieval Approach for Biometric Security using Color, Texture, and Shape Features Controlled by Fuzzy Heuristics. *J. of Computer and System Sciences* 78, 1258–1277 (2012)
8. Jain, A., Vailaya, A.: Image Retrieval using Color and Shape. *Pattern Recognition* 29(8), 1233–1244 (1996)
9. Kato, T.: Database Architecture for Content Based Image Retrieval, Image Storage and Retrieval Systems. In: *Proc. of SPIE*, vol. 1662, pp. 112–123 (1992)
10. Martinez, A.M., Benavente, R.: The AR Face Database. *CVC Report* 24 (1998)
11. Martinez, A.M., Kak, A.C.: PCA versus LDA. *IEEE Trans. on Pattern Anal. and Mach. Intell.* 23(2), 228–233 (2001)
12. Monwar, M., Gavrilova, M.L.: FES: A System for Combining Face, Ear and Signature Biometrics Using Rank Level Fusion. In: *ITNG 2008*, pp. 922–927 (2008)
13. Nandwal, A., Shrivastava, R., Vyas, K.: Face Recognition using Content Based Image Retrieval (CBIR) Technique. *VSRD-IJCSIT* 2(7), 605–611 (2012)
14. Porwik, P., Zyguła, J., Doroz, R., Proksa, R.: Biometric Recognition System Based on the Motion of the Human Body Gravity Centre Analysis Medical Knowledge Bases and Management. *J. of MIT* 15, 61–70 (2010)
15. Singha, M., Hemchandra, K.: Content Based Image Retrieval using Color and Texture. *SIPIJ* 3(1), 39–57 (2012)
16. Sultana, M., Uddin, M.S.: Efficient Trademark Retrieval using Weighted Image Features: A GPU Based Approach, 1st edn., August 4. VDM Verlag, Germany (August 4, 2011)
17. Swain, M.J., Ballard, D.H.: Color Indexing. *Comp. Vis.* 7, 11–32 (1991)
18. Vadivel, A., Majumdar, A. K., Sural, S.: Performance Comparison of Distance Metrics in Content Based Image Retrieval Applications. *Info. Tech.*, 159–164 (2003)
19. Wang, J.Z., Li, J., Wiederhold, G.: SIMPLIcity: Semantics Sensitive Integrated Matching for Picture Libraries. *IEEE TPAMI* 23(9), 947–963 (2001)
20. Wang, C., Gavrilova, M.L.: Delaunay Triangulation Algorithm for Fingerprint Matching. In: *ISVD 2006*, pp. 208–216 (2006)
21. Wecker, L., Samavati, F.F., Gavrilova, M.L.: Iris Synthesis: A Reverse Subdivision Application. In: *Graphite 2005*, pp. 121–125 (2005)
22. Yue, J., Li, Z., Liu, L., Fu, Z.: Content Based Image Retrieval using Color and Texture Fused Features. *Math. Comput. Modelling* 54, 1121–1127 (2011)

Face Recognition Based on Sequence of Images

Jacek Komorowski and Przemyslaw Rokita

Abstract. This paper presents a face recognition method based on a sequence of images. Face shape is reconstructed from images using a combination of structure-from-motion and multi-view stereo methods. The reconstructed 3D face model is compared against models held in a gallery. The novel element in the presented approach is the fact, that the reconstruction is based only on input images and doesn't require a generic, deformable face model. Experimental verification of the proposed method is also included.

1 Introduction

Three dimensional face recognition is an active and growing field of research [1] [2]. Using spatial information allows to mitigate some of the problems faced by methods based solely on visual information. 3D face recognition methods are less dependent on face pose and lighting variations. One of the barriers to a mass deployment of this technology is a difficulty with a face shape acquisition. Active vision techniques, such as laser scanning, are not appropriate for practical usage. Laser scanners are rather large, expensive and may be damaging to human eyes. Alternative, passive techniques, such as stereovision, multi-view stereo or structure-form-motion, are not very well suited for human face shape reconstruction. These methods are based on finding corresponding points on multiple images, that is points which are projections of the same scene point. Human skin has a relatively homogeneous texture which makes an automatic matching a difficult task.

Majority of methods which use passive vision techniques for face shape reconstruction, either uses complex image acquisition setup (e.g. set of 5 cameras [11]) or utilises a generic, deformable face models (e.g. [3]). Complex camera setups

Jacek Komorowski · Przemyslaw Rokita

Military University of Technology, Faculty of Cybernetics, Warsaw, Poland

e-mail: jac99@o2.pl, p.rokita@ii.pw.edu.pl

complicate practical deployment. Model-based approach is criticized [4], that it doesn't allow to model subtle details important for accurate face recognition, as reconstruction result are limited by a model parameter space.

The method presented in this paper uses a sequence of images from a single camera. Therefore it's easy to use as there's no need for a complicated equipment. Additionally it's based solely on input images and doesn't require a generic face model. Multi-view stereo algorithms can be used to reconstruct a 3D object model from a set or sequence of images taken from multiple viewpoint. Over the last years a significant progress was made in this area and a number of high-quality algorithms were developed. Best methods reviewed in [10] can deal with very demanding scenarios, where input images depict objects with little texture, containing few points which can be automatically matched across multiple images. For very demanding DinoRing¹ test set, containing images of a plaster dinosaur taken from multiple viewpoints, the best algorithms surveyed in [10] were able to reconstruct over 90% of the object surface with error below 0.4 mm. Unfortunately multi-view stereo algorithms assume that all images are fully calibrated, that is both intrinsic (camera focal length, distortion coefficients) and extrinsic (camera pose) parameters for each image are known. Such algorithms cannot be used when a sequence contains images of an object moving freely in front of the camera. Intrinsic camera parameters are fixed, and can be estimated with a prior calibration. But extrinsic parameters are different for each image and cannot be easily estimated. To use some high-quality multiview-stereo algorithm for face shape reconstruction from a sequence of images, extrinsic parameters for each image in the sequence must be estimated.

2 Details of the Method

This section describes details of our face recognition method. The method is based on a sequence of images from a monocular camera. It's assumed that a person sits in front of the camera and is asked to rotate his head left and right. An exemplary input sequence is depicted on Fig. 1.



Fig. 1 Exemplary input sequence (5 from 80 images)

¹ <http://vision.middlebury.edu/mview/data>

Main steps of the presented method are depicted on Fig. 2.

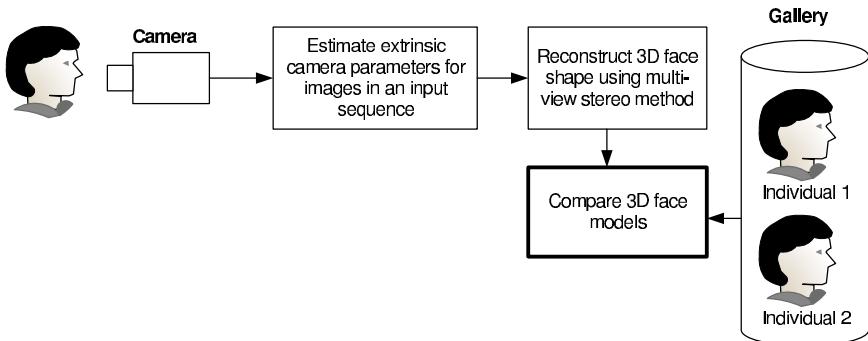


Fig. 2 Recognition system concept

Step 1

Extrinsic camera parameters (rotation matrix \mathbf{R} and translation vector \mathbf{T}) are estimated for each image in the sequence. This is done using a method developed by authors and described in [7] and [8]. The method is designed to work well with demanding scenarios, where input images contain little texture. It doesn't use a generic, deformable face model and is based solely on input data. Results of this step are depicted on Fig. 3.

Step 2

Once camera extrinsic parameters are estimated, any multi-view stereo algorithm can be used to reconstruct a 3D face shape. In our implementation a patch-based multi-view stereo method PMVS [5]² was used. An input to the PMVS algorithm is a sequence of images and estimated camera extrinsic parameters. The output is a cloud of oriented points (see Fig. 4).

Step 3

Face model reconstructed from an image sequence is compared with models in a gallery. Distance between point clouds is used as a similarity measure between two face models. Distance between two point clouds is defined as an average Euclidean distance between each point from the first model to the closest point in the second model. Two face models usually do not fully overlap. Due to differences in input

² <http://grail.cs.washington.edu/software/pmv/>

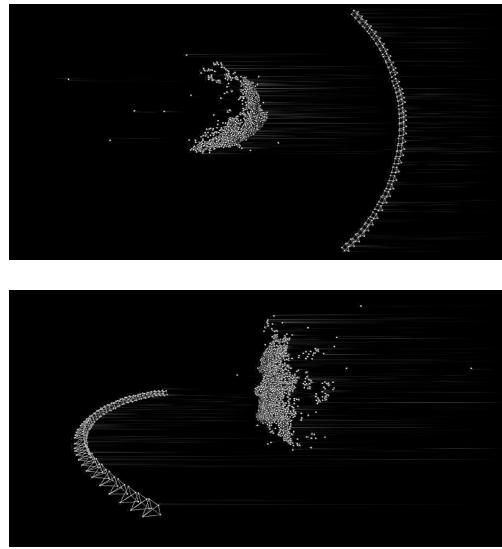


Fig. 3 Estimated camera poses (pyramids) and a sparse face model (point cloud) based on a sequence of images from Fig. 1

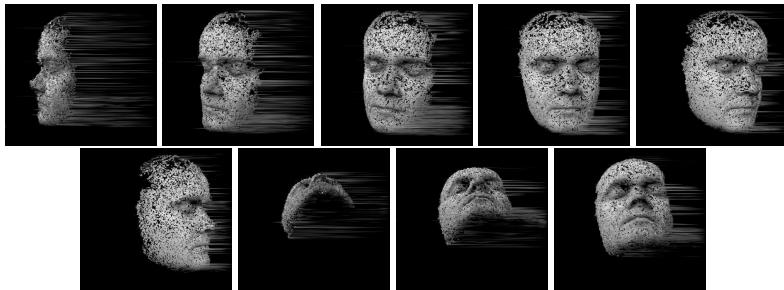


Fig. 4 Face reconstruction results based on a sequence from Fig. 1

sequences³ one cloud may contain regions from a reconstructed object surface, not presented in the second model. To deal with this problem a relatively simple heuristic is used. A median distance between each point from the first model and the closest point in the second model is calculated, and points further away than some small multiple of the median are discarded. Formal definition of the distance metric used to compare 2 point clouds is as follows:

Let $\mathcal{C}_1 \subset \mathbb{R}^3$ and $\mathcal{C}_2 \subset \mathbb{R}^3$ denote two clouds consisting of points in 3D Cartesian space. $d(p, \mathcal{C})$ denotes a distance of a point $p \in \mathbb{R}^3$ from the cloud $\mathcal{C} \subset \mathbb{R}^3$, defined as:

³ E.g. different maximum face rotation angle.

$$d(p, \mathcal{C}) = \min_{p' \in \mathcal{C}} \|p' - p\| , \quad (1)$$

where $\|p' - p\|$ is an Euclidean distance between points p i p' . Distance between point cloud \mathcal{C}_1 and \mathcal{C}_2 with a threshold k is defined as:

$$d_k(\mathcal{C}_1, \mathcal{C}_2) = \frac{1}{|\mathcal{C}_1 \setminus \mathcal{O}_k|} \sum_{p \in \mathcal{C}_1 \setminus \mathcal{O}_k} d(p, \mathcal{C}_2) , \quad (2)$$

where \mathcal{O}_k is a set of points in a cloud \mathcal{C}_1 not having close neighbours in a cloud \mathcal{C}_2 , defined as:

$$\mathcal{O}_k = \{p \in \mathcal{C}_1 \mid d(p, \mathcal{C}_2) > km\} , \quad (3)$$

where m is a median distance between each point from the first cloud and the closest point from the second cloud. In the implementation of the presented method threshold $k = 4$ was chosen.

Two face models being compared may have a different scale and orientation. Scale difference is caused by the fact, that extrinsic parameters can be estimated for a sequence of images only up to an unknown scale factor. Thus a metric reconstruction is also possible up to a scale factor. Orientation may be different because reconstructed head pose is aligned with the head pose on the first image. In order to calculate a distance between two point clouds, they must be aligned first. We use a variant of the popular ICP⁴ [9] algorithm, which can find a rigid body transformation aligning two point clouds.

Let C_s denotes a source point cloud and C_d a destination point cloud. Our modified version of ICP method has the following steps:

1. Compute centroids of a source and destination cloud
 - a. $\bar{c}_s = (\sum_{p \in \mathcal{C}_s} p) / |\mathcal{C}_s|$
 - b. $\bar{c}_d = (\sum_{q \in \mathcal{C}_d} q) / |\mathcal{C}_d|$
2. Scale a source point cloud to match a destination cloud scale using a formula from [6]:
 - a. Compute scaling factor: $scale = \sqrt{\frac{\sum_{q \in \mathcal{C}_d} \|q_i - \bar{c}_d\|^2}{\sum_{p \in \mathcal{C}_s} \|p_i - \bar{c}_s\|^2}}$
 - b. Multiply coordinates of points in \mathcal{C}_s by $scale$
3. Align centroid of a source point cloud with a centroid of a destination cloud
 - a. Translate all point in C_s by a vector $\bar{c}_d - \bar{c}_s$.
4. Choose a random sample $\mathcal{S} = \{p_i\}$ of s points from a source cloud C_s
5. Match each point from a sample S with the closest point in a destination cloud C_d . Let $\mathcal{M} = \{(p_i, q_i)\}$ denotes a set of corresponding point.

⁴ ang. Iterative Closest Point.

6. Remove outliers from \mathcal{M} , that is remove pairs (p_i, q_i) for which $|d_i - q_i| > km$, where m is a median distance between pairs of corresponding points in \mathcal{M} , and k is a small integer⁵.
7. Find a rigid body transformation (rotation matrix \mathbf{R} and translation vector \mathbf{T}) minimizing error metric $E(\mathbf{R}, \mathbf{T})$ and apply the transformation on a source point cloud \mathcal{C}_s
8. If number of iterations $< N$, go to point 4 else terminate the algorithm

Algorithm parametrization and error metric E were chosen experimentally to achieve good convergence and a reasonable running time. Sample size M is set to 500 (out of app. 40'000 points in clouds) and number of iterations $N = 15$, as it was verified that larger values increase running time but do not improve convergence. As an error metric E , a point-to-plane error metric is chosen as it gives much faster convergence than a classic point-to-point error metric. Point-to-plane error metric is given by the formula [9]:

$$E_{\text{point-to-plane}}(\mathbf{R}, \mathbf{T}) = \sum_i ((\mathbf{R}p_i + \mathbf{T} - q_i) \cdot n_i)^2 , \quad (4)$$

where n_i is a normal to the destination cloud surface at point q_i .

3 Experiments

This section presents results of an experimental verification of accuracy of the face recognition method presented in this paper. Test database built by authors contains 81 image sequences of 27 individuals, 3 sequences per one person. Images were acquired with Point Grey Chameleon camera⁶ with 800x600 pixels resolution. In each sequence a persons sitting in front of a camera is asked to rotate his head right and left. Exemplary sequences are depicted on Fig. 5. The database was split into 2 parts: 27 image sequences (1 per each individual) were used to build a gallery, 54 sequences (2 per each individual) were used to build a test set.

Error metrics

Face recognition system can be used to perform 2 tasks: verification and identification. Verification is a task where the biometric system attempts to confirm an individual's claimed identity. 2 error metrics are used to assess accuracy of an identity verification task: FAR⁷ and FRR⁸. FAR is defined as a ratio of a number of attempts when an identity was falsely positively verified to a number of all attempts. FRR is

⁵ In implementation $k = 4$ was chosen.

⁶ http://www.ptgrey.com/products/chameleon/chameleon_usb_camera.asp

⁷ False Acceptance Ratio.

⁸ False Rejection Ratio.



Fig. 5 Exemplary sequences from a test database

defined as a ratio of a number of attempts when an identity was falsely negatively verified to a number of all attempts.

Identification is a task where biometric system searches a gallery for a reference matching submitted biometric sample, and if found, returns a corresponding identity. Accuracy of identification tasks is measured with a CMC⁹ curve. CMC is a function of a recognition rate as a number of best n -matches considered. For a given n , recognition rate is a ratio of attempts when a chosen individual from a test set was among n closest matches in the gallery to number of all attempts. Clearly, when n equals to the number of individuals in the gallery, recognition rate is equal to one.

Experiment 1

In this experiment accuracy of identity verification scenario was tested. Each sequence from a test set was used to reconstruct a 3D face model which was matched against each face model in the gallery. If the distance between face model from a test set and a face model from a gallery was below a threshold Θ the identity was positively verified. Otherwise identity was negatively verified.

Both FAR and FRR are dependent on threshold Θ . When it's increased, more distant faces are identified as belonging to the same individual thus leading to FAR

⁹ Cummulative Match Characteristics.

increase and FRR decrease. Fig. 6 depicts values of FAR as a function of a threshold Θ . Fig. 7 shows values of FRR as a function of a threshold Θ .

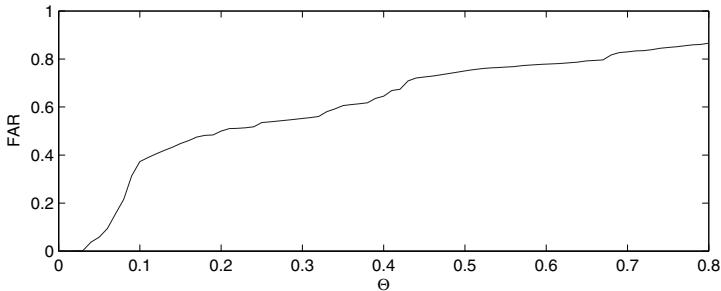


Fig. 6 FAR as a function of a threshold Θ

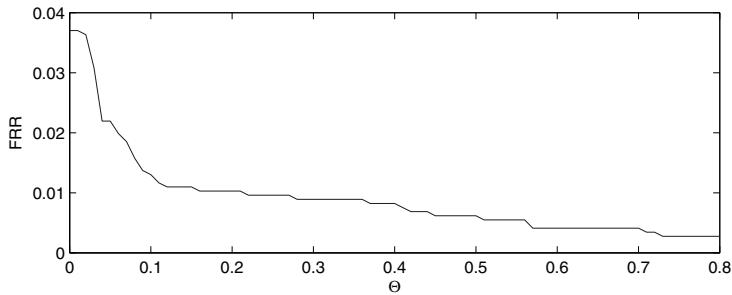


Fig. 7 FRR as a function of a threshold Θ

The trade-off between FAR and FRR rates is expressed using ROC¹⁰ curve and is shown on Fig. 8. ERR¹¹, that is a rate at which FAR = FRR is equal to 0.025 and is a rather low value. It means that in 2.5% of attempts identity was falsely positively verified and in 2.5% of attempts identity was falsely negatively verified.

Experiment 2

In this experiment identification in a closed-set scenario was tested, as each individual from a test set was present in the gallery. Each sequence from a test set was used to reconstruct a 3D face model which was matched against each face model in the gallery. Models with the closest distance were declared as a match.

¹⁰ Receiver Operating Characteristic.

¹¹ Equal Error Rate.

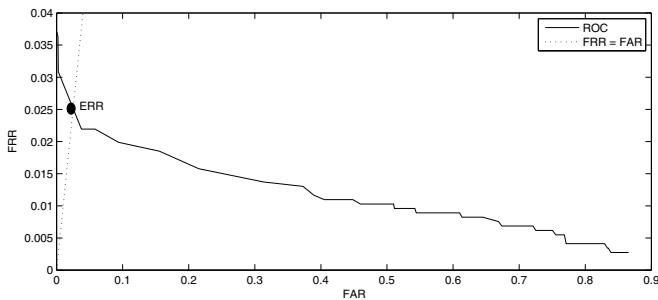


Fig. 8 ROC curve and ERR point

Fig. 9 shows resultant CMC curve. When finding a single, best match in the gallery ($n = 1$) for each individual from a test set, the method achieved almost 75% accuracy. If considering 5 best matches in the gallery ($n = 5$), over 90% accuracy was achieved.

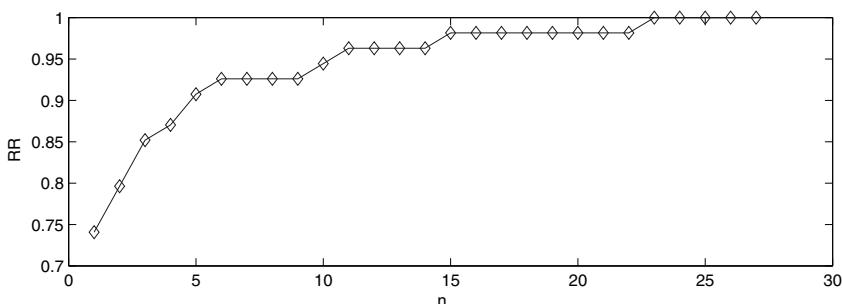


Fig. 9 CMC curve

4 Conclusions and Future Work

The presented method allows to achieve a reasonably good face recognition accuracy. Although the results should be taken with care, as they were obtained using a relative small test database. To ensure validity of the proposed approach, experiments using much larger test database should be done. For face recognition a relatively simple approach is used, based on direct comparison of two point clouds. It's worth to investigate more advanced approaches, e.g. based on comparison of a local characteristics such as nose profile, or a relative position of eyes, nose and lips. It must be noted that recognition is based only on spatial information and 2D information (texture) is not used. Combining 2 modalities (shape and texture) may allow to achieve better recognition rates.

Face reconstruction method presented in this paper consists of two separate and distinct steps. Sparse point cloud build during the process of estimating extrinsic parameters is discarded, and only extrinsic parameters are passed to the second step (multi view stereo reconstruction). Potentially 3D points from a sparse point cloud created in the first step can be used to initialise multi-view stereo reconstruction process.

References

1. Bowyer, K., Chang, K., Flynn, P.: A survey of 3d and multi-modal 3d+2d face recognition. *Computer Vision and Image Understanding* 101(1), 1–15 (2006)
2. Chang, K., Bowyer, K., Flynn, P.: An evaluation of multi-modal 2d+3d face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 619–624 (2005)
3. Cheng, C., Lai, S.: An integrated approach to 3d face model reconstruction from video. In: *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems* (2001)
4. Fidaleo, D., Medioni, G.: Model-assisted 3d face reconstruction from video. In: *Proceedings of the 3rd International Conference on Analysis and Modeling of Faces and Gestures* (2007)
5. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(8), 1362–1376 (2010)
6. Horn, B.: Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America* 4(4), 629–642 (1987)
7. Komorowski, J., Rokita, P.: Camera pose estimation from sequence of calibrated images. *Advances in Intelligent Systems and Computing, Image Processing and Communication Challenges* 4 184, 101–110 (2012)
8. Komorowski, J., Rokita, P.: Extrinsic camera calibration method and its performance evaluation. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) *ICCVG 2012. LNCS*, vol. 7594, pp. 129–138. Springer, Heidelberg (2012)
9. Rusinkiewicz, S., Levoy, M.: Efficient variants of the icp algorithm. In: *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling*, pp. 145–152 (2001)
10. Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2006)
11. Spreeuwiers, L.: Multi-view passive acquisition device for 3d face recognition. *Proceedings of the Special Interest Group on Biometrics and Electronic Signatures* 137, 13–24 (2008)

Statistical Analysis in Signature Recognition System Based on Levenshtein Distance

Malgorzata Palys, Rafal Doroz, and Piotr Porwik

Abstract. In this paper we develop our previously presented studies, where adaptation of the Levenshtein method in a signature recognition process is proposed. Three methods based on the normalized Levenshtein measure were taken into consideration. The studies included an analysis and selection of appropriate signature features, on the basis of which the authenticity of a signature was verified later. A statistical apparatus was used to perform a comprehensive analysis. Results obtained were tested by means of χ^2 independence test. It allowed determining the relationship between signature features and the errors of classifier.

1 Introduction

In the modern world security problems are increasingly very important, because safety of goods, resources and data should be protected. In order to protect them, common methods based on human knowledge are used, for example: passwords and PIN codes, as well as methods based on identifiers, e.g. identity cards and credit cards. These methods may not be able to serve their purpose for various reasons, such as forgetting a password or a PIN code, giving it to another person, or identifier loss, theft or forgery. In the era of computerization and automation, the gap in the problems related to protections is filled by biometric techniques. One of the most popular biometric techniques is a handwritten signature. Signature verification from biometric features point of view presents some advantages, such as: non invasive, intuitive and fast, well accepted socially and legally. Additionally, signature verification generally has a low storage requirements. The effectiveness of the use of an analysis of handwritten signatures as a biometric technique is very high. The main

Malgorzata Palys · Rafal Doroz · Piotr Porwik
Institute of Computer Science, University of Silesia, ul. Bedzinska 39, 41-200 Sosnowiec
e-mail: {malgorzata.palys, rafal.doroz, piotr.porwik}@us.edu.pl

factor affecting the effectiveness is selection of an appropriate signature recognition method. Currently a lot of different approaches have been proposed for signature verification in the literature [2], [3], [6], [7]. It should be noted that verification and identification of objects is also proposed in other scientific problems [4], [8].

This study presents a method of comparing signatures with the use of the normalized Levenshtein metrics [9], [13]. The effectiveness of these metrics in the process of signature recognition has been examined. A large number of results was obtained, which made an analysis more difficult. Therefore, the presented method includes a detailed statistical analysis, which allows approach to select features of the signatures being adequately compared.

2 Feature Preparation

Biometrics signature is one of the longest-known security techniques invented by humans. Signature has for many years adopted a form of determining the credibility (such as during operations related to running a bank account). Data collection process within a signature recognition process can be divided into two categories: static and dynamic. The static system collects data using *off-line* devices. A signature is put on paper and then is converted into a digital form with the use of a scanner or a digital camera. In this case, the shape of the signature is the only data source, without the possibility of using dynamic data. Signature recognition on the basis of photos does not protect against fraud. On the other hand, dynamic systems use *on-line* devices, which register, apart from the image of the signature, also dynamic data connected with it. The most popular *on-line* devices are graphics tablets. By using tablet, a signature can be recorded in the form of an n -point set. Fig 1. presents an example of signature S_i , captured by tablet.

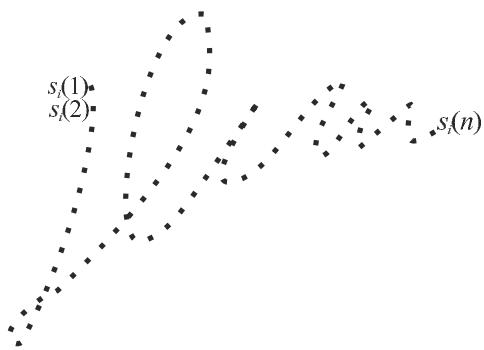


Fig. 1 An example of signature S_i

Signature S_i can be represented by the vector of points:

$$S_i = \begin{bmatrix} s_i(1) \\ s_i(2) \\ \vdots \\ s_i(n) \end{bmatrix}, \quad (1)$$

where:

$s_i(j)$ – the j -th point of signature S_i .

The tablet, during signing, is capable to measure many dynamics parameters, such as a pressure pen on tablet surface, position of pen, the angles at which the person holds a pen. This implies that there is a dynamic feature vector $c_i(u)$ associated with each point $s_i(u)$ of signature S_i :

$$s_i(u) \rightarrow c_i(u). \quad (2)$$

The local position (x_i, y_i) of the pen is given directly by a graphics tablet, while speed and acceleration can be obtained from this device or can be calculated on the basis of the position parameter. In presented work, following signature features were used:

$X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$ - sets of coordinates of n signature points,
 $P = \{p_1, p_2, \dots, p_n\}$ - set of pressure in particular points of signature,

Having the sets of mentioned features it is possible to determine, basing on them, additional features like as:

$V = \{v_1, v_2, \dots, v_n\}$ - set of the speed of the pen in successive signature points [3],
 $V_{up} = \{v_{up1}, v_{up2}, \dots, v_{upn}\}$ - set of the positive velocity values of the pen in successive signature points [5],

$V_{down} = \{v_{down1}, v_{down2}, \dots, v_{downn}\}$ - set of the negative velocity values of the pen in successive signature points [5],

$V_x = \{v_{x1}, v_{x2}, \dots, v_{xn}\}$ - set of the horizontal speed of the pen in successive signature points [3],

$V_y = \{v_{y1}, v_{y2}, \dots, v_{yn}\}$ - set of the vertical speed of the pen in successive signature points [3],

$P_{ch} = \{p_{ch1}, p_{ch2}, \dots, p_{chn}\}$ - set of the change in the pen pressure in successive signature points [5],

$K = \{k_1, k_2, \dots, k_n\}$ - set of the inverse of the radius of the curve in successive signature points [12].

Because in proposed approach 10 signature features have been selected the extracted features form the vector:

$$c_i(u) = [c_i^1(u), c_i^2(u), \dots, c_i^{10}(u)], \quad (3)$$

namely:

$$c_i(u) = [x_i(u), y_i(u), \dots, k_i(u)]. \quad (4)$$

Finally, the signature can be described by the matrix S_i :

$$S_i = \begin{bmatrix} x_i(1) & y_i(1) & \cdots & k_i(1) \\ x_i(2) & y_i(2) & \cdots & k_i(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_i(n) & y_i(n) & \cdots & k_i(n) \end{bmatrix}. \quad (5)$$

3 Normalization of the Levenshtein Distance

The Levenshtein distance is the number of certain operations, called elementary operations, which must be performed to transform one character string into another one [2], [13].

Let Σ define an alphabet of characters and a set containing all character sub-strings from this alphabet Σ' . Then, let's define two character strings and belonging to Σ' , where n and m are the lengths of these strings. Let $T_{A,B} = T_1, T_2, \dots, T_l$ stand for the transformation of A character string into B character string with the use of the finite number l of elementary operations. Elementary operations are performed on the pair of characters (a, b) , where $a, b \neq \lambda$ described more often as $(a \rightarrow b)$. The sign λ represents an empty character, which does not belong to the alphabet. Three elementary operations can be distinguished:

- D – deleting a character $(a \rightarrow \lambda), (b \rightarrow \lambda)$,
- I – inserting a character $(\lambda \rightarrow a), (\lambda \rightarrow b)$,
- R – replacing a character $(a \rightarrow b), (b \rightarrow a)$.

Each elementary operation has a specific cost of its performance, which is called a weight of a given elementary operation. The weighting function δ assigns a non-negative real number to the i -th elementary operation $(a \rightarrow b)$:

$$\delta(T_i) = \delta(a \rightarrow b). \quad (6)$$

The weight of the $T_{A,B}$ transformation can be calculated using the following formula:

$$\delta(T_{A,B}) = \sum_{i=1}^l \delta(T_i). \quad (7)$$

The $T_{A,B}$ transformation can be defined for a specific path of transition from the A character string into the B character string. Let the $P_{A,B} = \{P_{A,B}^1, P_{A,B}^2, \dots, P_{A,B}^h\}$ set contain all possible paths of transitions from the A character string into the B character string, where h is the number of all possible transition paths.

Let $W(P_{A,B})$ be a function calculating weights of individual paths from the $P_{A,B}$ set:

$$W(P_{A,B}) = \delta(T_{A,B}). \quad (8)$$

The General Levenshtein Distance (GLD) for the two character strings A, B being compared can be defined as follows:

$$GLD(A, B) = \min \{ \delta(T_{A,B}) \} = \min \{ W(P_{A,B}) \}. \quad (9)$$

As the final value of the Levenshtein distance calculated for two character strings is included in the $[0, \infty)$ interval, it is not possible on this basis to determine the percentage similarity of the strings being compared. This considerably hinders the evaluation of similarity of the strings being compared. *Ned1* metric is defined by the formula:

$$Ned1(A, B) = \min \left\{ \frac{W(P_{A,B})}{|P_{A,B}|} \right\}, \quad (10)$$

where:

$|P_{A,B}|$ – the number of elementary operations in an individual path. Another proposed measure is the *Ned2* metric described by the following formula:

$$Ned2(A, B) = \min \left\{ \frac{W(P_{A,B})}{|A| + |B|} \right\} = \frac{GLD(A, B)}{|A| + |B|}, \quad (11)$$

where:

$|A| + |B|$ – is the sum of lengths of the A and B strings.

Third modification of the Levenshtein distance, used in this study, is the d_{N-GLD} distance. This distance is expressed by the formula:

$$d_{N-GLD}(A, B) = \frac{2 \cdot GLD(A, B)}{\max(D, I) \cdot (|A| + |B|) + GLD(A, B)}, \quad (12)$$

where:

D – the cost of deleting a character,

I – the cost of inserting a character.

All presented metrics: *Ned1*, *Ned2*, d_{N-GLD} return results from the $[0, 1]$ interval. If two strings being compared are the same, the metrics return the 0 value. For further assessment of their effectiveness with the use of EER coefficient, the metrics (13), (14) and (15) were adequately modified, so that the result of the comparison of two identical strings was the value 1:

$$NED1(A, B) = 1 - Ned1(A, B), \quad (13)$$

$$NED2(A, B) = 1 - Ned2(A, B), \quad (14)$$

$$NGLD(A, B) = 1 - d_{N-GLD}(A, B). \quad (15)$$

4 The Use of Normalized Levenshtein Metrics in the Process of Recognition of Handwritten Signatures

Levenshtein distance calculates the distances between two strings. In the presented method the values of individual feature of signatures S_i and S_j are compared. In order to use of normalized Levenshtein metrics in the process of recognition of handwritten signatures, new signature similarity measure LM_{ij}^t has been introduced. This measure is based on similarity metrics between two features t signature being compared. It will be described on the example of the *NED1* metric:

$$LM_{ij}^t = NED1(S_i(t), S_j(t)), \quad (16)$$

where:

$S_i(t)$ – t -th column (feature) of signature matrix S_i .

At the beginning the values of signature features were normalized to the [0,1] range, so they can take indefinitely many values from this range [1], [10]. Therefore, the probability of occurrence of two identical feature values in two compared strings is near zero. In order to eliminate this situation, the additional parameter ϑ was introduced. It determines, to what maximum extent the two values being compared can differ from each other in order to be treated as equal. The features values $c_i^r(u) = c_j^r(u)$, if it fulfils the following condition:

$$|c_i^r(u) - c_j^r(u)| < \vartheta, \quad (17)$$

where:

ϑ – the maximum difference between the values of the features that allows recognizing them as equal,

$c_i^r(u)$ – the u -th element of r -th feature of the signature S_i ,

$c_j^r(u)$ – the u -th element of r -th feature of the signature S_j .

The evaluation of the similarity of individual signatures was performed on the basis of an analysis of ten signature features and their combination. In order to specify the influence of a given feature on the result of the comparison, the weights w_1 , w_2 , w_3 of the features were introduced. Thus, 35 different values were obtained as the result of the comparison, and each of them described the similarity of a different combination of signature feature. The formula for determining the $WLM(S_i, S_j)$ similarity value of the two signatures S_i and S_j , taking into account combination of signature feature, is as follows:

$$WLM(S_i, S_j) = \left\{ LM_{ij}(w_1, w_2, w_3) : w_1, w_2, w_3 \in N \wedge \sum_{i=1}^3 w_i = 1 \right\}, \quad (18)$$

where:

$$\forall \begin{array}{l} LM_{ij}(w_1, w_2, w_3) = LM_{ij}^k \cdot w_1 + LM_{ij}^l \cdot w_2 + LM_{ij}^m \cdot w_3 \\ k, l, m \in F \\ k \neq l \neq m \end{array} \quad (19)$$

N - weight of the signature feature, $N = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$,

F - number of feature, $F = \{1, 2, \dots, 10\}$.

The individual element of the set F correspond to the number of columns of matrix S_i . For example, $F = 1$ means the feature X .

5 The Course and Results of the Studies

The studies were conducted for 50 signatures coming from different persons. The set of test signatures used in the studies comes from the SVC2004 database. The signatures were divided into 10 groups. Each group contained 4 original signatures of one person and 1 forged signature. In order to assess, which combination of signature feature has the greatest impact on EER values, the χ^2 test was applied. It allows determining whether there is a relationship between feature combinations and EER values.

In order to perform the χ^2 test, two hypotheses should be made: H_0 and H_1 . The null hypothesis H_0 assumes that selection of features does not affect the effectiveness of signature comparison using the Levenshtein method:

$$H_0 : P(Z = z_k \cdot U = u_m) = P(Z = z_k) \cdot P(U = u_m). \quad (20)$$

The alternative hypothesis H_1 shows a relationship between the Z and U :

$$H_1 : P(Z = z_k \cdot U = u_m) \neq P(Z = z_k) \cdot P(U = u_m), \quad (21)$$

where the variable Z is a combination of the signature features:

$$Z = \{X, Y, P, V, V_{up}, V_{down}, V_x, V_y, P_{ch}, K, XY, XP, XV, X V_{up}, X V_{down}, \\ X V_y, X V_x, X P_{ch}, X K, YP, YV, Y V_{up}, Y V_{down}, Y V_y, Y V_x, Y P_{ch}, Y K, \\ Y YP, X YV, X Y V_{up}, X Y V_{down}, X Y V_y, X Y V_x, X Y P_{ch}, X Y K\}.$$

The variable U is a range of EER values. As the number of results for each of the three analysed measures was very high (1736733), the analysed data were divided into 7 subsets. Each subset was assigned to a different EER range. Boundaries of division are determined by dividing the range between the highest and lowest value into 7 equal parts. Each range was named depending on the value of the errors it contained. For example, for the *NED1* measure (in which the lowest value of EER = 1.161%, and the highest value of EER = 54.918%), the determined ranges are presented in Table 1 .

Table 1 Table of ranges of EER values determined for the *NED1* measure

Name of range Range EER [%]	
Excellent	[1.161-8.841)
Very good	[8.841-16.521)
Good	[16.521-24.201)
Average	[24.201-31.881)
Poor	[31.881-39.561)
Bad	[39.561-47.241)
Very bad	[47.241-54.921)

Basing on the assumptions presented in Table 1 , the quantity table was prepared, which contains the quantity of EER values obtained for different combinations of signature features. Then the expected quantities was calculated [11]. The seven feature combinations of the largest differences calculated between actual quantities and expected quantities were presented in Table 2 .

Table 2 Table showing the difference between the actual quantities and expected quantities of EER values for the *NED1* measure

	<i>XY</i>	<i>XV</i>	<i>XP_{ch}</i>	<i>XYP</i>	<i>XYV_{up}</i>	<i>XYV_y</i>	<i>XYV_x</i>
Excellent	1366.90	-1033.10	-1038.83	3611.10	1265.38	1008.65	1931.93
Very good	64.54	-139.46	-501.12	80.22	342.56	-5.10	82.25
Good	-642.35	440.65	93.87	-1530.14	-619.92	-618.69	-770.47
Average	-488.51	534.49	596.59	-1227.37	-529.27	-338.17	-720.07
Poor	-228.56	170.44	634.48	-691.33	-331.29	-90.24	-413.20
Bad	-69.06	28.94	210.95	-232.12	-126.11	36.90	-107.08
Very bad	-2.95	-1.95	4.05	-10.36	-1.36	6.65	-3.35

For the *NED1* measure, the calculated statistic is $\chi^2 = 32356.1$. The critical value $\chi_{\alpha}^2 = 238.32$ was taken from the distribution tables χ^2 for the adopted level of significance $\alpha = 0.05$. The quantity table has 7 rows and 35 columns, so $s = (7 - 1)(35 - 1) = 204$ degrees of freedom. The calculated statistic belongs to the critical area ($\chi^2 > \chi_{\alpha}^2$). Therefore the null hypothesis should be rejected in favour of the alternative hypothesis that assumes that these combinations affect the range of EER values. In addition, basing on Table 2 , it can be stated that the greatest impact on the EER value in the Levenshtein method has a combination of *XYP* features, and therefore the use of this combination will allow increasing the effectiveness of signature comparison by this method.

A similar analysis was carried out for the *NED2* and *NGLD* measures. Table 3 showing the difference between the actual quantities and expected quantities of EER values for the *NED2* measure whereas Table 4 for the *NGLD* measure.

Table 3 Table showing the difference between the actual quantities and expected quantities of EER values for the *NED2* measure

	<i>XY</i>	<i>XV_{down}</i>	<i>XV_y</i>	<i>XK</i>	<i>XYP</i>	<i>XYV_{up}</i>	<i>XYV_x</i>
Excellent	1749.63	-1199.49	-1218.34	-1199.14	4622.21	1619.68	2472.87
Very good	86.48	168.22	-67.16	168.68	97.86	417.92	100.34
Good	-610.23	546.87	249.73	547.08	-1912.67	-774.90	-963.09
Average	-561.78	172.60	242.39	171.87	-1595.58	-688.05	-936.09
Poor	-205.71	53.44	442.82	53.52	-394.06	-188.83	-235.52
Bad	-110.50	-8.10	205.78	24.87	-336.58	-182.86	-155.27
Very bad	-2.39	-0.77	10.57	-0.80	-8.70	-1.14	-2.82

Table 4 Table showing the difference between the actual quantities and expected quantities of EER values for the *NGLD* measure

	<i>XY</i>	<i>XV_{up}</i>	<i>XP_{ch}</i>	<i>YP_{ch}</i>	<i>XYP</i>	<i>XYV_{up}</i>	<i>XYV_x</i>
Excellent	2152.04	-1333.68	-1595.64	-1335.05	5546.65	1943.62	2967.44
Very good	76.97	116.32	-765.51	-810.91	111.56	476.43	114.39
Good	-488.19	213.29	71.34	32.92	-1472.76	-960.87	-1194.23
Average	-449.43	202.96	425.61	796.12	-1818.96	-784.38	-1067.14
Poor	-242.73	275.70	1272.77	434.33	-464.99	-171.84	-214.33
Bad	-114.92	68.12	571.41	607.07	-400.53	-217.60	-184.77
Very bad	-2.49	5.09	3.41	2.38	-6.09	-1.31	-3.24

Statistics for the *NED2* and *NGLD* measures are respectively $\chi^2 = 96432.7$ and $\chi^2 = 78784.6$. Thus, they belong to the same critical area as the *NED1* measure. Similarly as the *NED1* measure, it has been found that the *XYP* feature had the greatest impact on the EER value in signature recognition with the use of the Levenshtein method.

6 Conclusions

In this paper the method of feature selection with statistical significance testing was proposed. The study focused on determining a combination of dynamic features of signatures which allows obtaining the lowest error in signature recognition. The analysis proves that there is a statistical relationship between signature features and

the error returned by the classifier based on the normalized Levenshtein method. From obtained results follow that the best features selection is given by combination of feature XYP . For these parameters the EER coefficient achieves the lowest values. In the future the result obtained by means of the test χ^2 will be compared with other tests known from the literature. Also other features of signatures will be taken into account.

References

1. Cha, S.: Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences* 1(4), 300–307 (2007)
2. Doroz, R., Porwik, P.: Handwritten signature recognition with adaptive selection of behavioral features. In: Chaki, N., Cortesi, A. (eds.) CISIM 2011. CCIS, vol. 245, pp. 128–136. Springer, Heidelberg (2011)
3. Doroz, R., Wróbel, K.: Method of signature recognition with the use of the mean differences. In: Proceedings of the ITI 2009 31st International Conference (ITI 2009), pp. 231–235 (2009)
4. Froelich, W., Wakulicz-Deja, A.: Probabilistic Similarity-Based Reduct. In: Yao, J., Ramaulla, S., Wang, G., Suraj, Z. (eds.) RSKT 2011. LNCS, vol. 6954, pp. 610–615. Springer, Heidelberg (2011)
5. Gupta, G.K.: The state of the art in on-line handwritten signature verification4 (2006)
6. Impedovo, S., Pirlo, G.: Verification of handwritten signatures: an overview. In: 14th International Conference on Image Analysis and Processing (ICIAP 2007), pp. 191–196 (2007)
7. Khan, M.K., Khan, M.A., Khan, M.A.U., Ahmad, I.: On-line signature verification by exploiting inter-feature dependencies. In: 18th International Conference on Pattern Recognition (ICPR 2006), vol. 2, pp. 796–799 (2006)
8. Koprowski, R., Wrobel: The cell structures segmentation. In: 4th International Conference on Computer Recognition Systems (CORES 2005), pp. 569–576 (2005)
9. Levenshtein, V.I.: Binary codes capable of correcting deletions, Insertions, And Reversals. In: Soviet Physics Dokl, pp. 707–710 (1966)
10. Marzal, A., Vidal, E.: Computation of normalized edit distance and applications. *IEEE Trans. Pattern Analysis and Machine Intelligence* 15(9), 926–932 (1993)
11. Para, T., Mitas, M.: Determining signatures characteristic features using statistical methods. *Journal of Medical Informatics and Technologies* 1, 41–50 (2008)
12. Pastor, M., Toselli, A., Vidal, E.: Writing speed normalization for on-line handwritten text recognition. In: Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition, pp. 1131–1135 (2005)
13. Schimke, S., Vielhauer, C., Dittmann, J.: Using adapted Levenshtein distance for on-line signature authentication. In: Proceedings of the 17th International Conference, vol. 2, pp. 931–934 (2004)
14. Weigel, A., Fein, F.: Normalizing the weighted edit distance. In: Proc. 12th IAPR Intl Conf. Pattern Recognition, Conf. B: Computer Vision and Image Processing., vol. 2, pp. 399–402 (1994)

A Computational Assessment of a Blood Vessel's Roughness

Tomasz Wesolowski and Krzysztof Wrobel

Abstract. High risk of heart attack due to the accumulation of atherosclerotic plaque in coronary arteries is currently a serious social problem. Therefore, it is obvious that current preventive measures include early detection of atherosclerosis process. Dedicated tests that allow to diagnose this disease are not commonly performed so specialists are trying to detect it in other ways. One of them is the manual analysis of computed tomography (CT) images in order to determine roughness of the coronary arteries. The paper presents a computer aided method for measuring the roughness of coronary vessels based on the analysis of digital CT images.

1 Introduction

Heart diseases all over the world are reasons of cardiovascular mortality. This kind of disease is announced by World Health Organization (WHO) as a serious civilisation problem. The most common reason of heart attack or stroke is atherosclerosis. The course of this disease depends on many factors which can be observed during medical examinations.

If the process of atherosclerosis is located within coronary arteries then a disease known as coronary artery disease (CAD) develops. Heart muscle is a braided network of vessels arising from the left and right coronary artery surrounding the heart. The network of vessels supplies blood, oxygen and nutrients. Normally, a balance is ensured between the blood flow to the heart and its needs, which increase significantly when we exercise. The balance is disrupted by disturbances in the transport. They are most often caused by the atherosclerosis leads to reduction in a cross-section of one of the coronary arteries. Some heart vessels changes can be observed by means of medical imaging techniques. Imaging technique permits for a simple

Tomasz Wesolowski · Krzysztof Wrobel

Institute of Computer Science, University of Silesia, ul. Bedzinska 39, 41-200 Sosnowiec
e-mail: {tomasz.wesolowski, krzysztof.wrobel}@us.edu.pl

visual assessment. From a medical point of view it is insufficient - if it is possible, all subjective assessments should be replaced by reliable and repeatable measurements. These measures are difficult to estimate because vessels' shape and topology are substantially different for patients examined.

Recent advances in non-invasive imaging of coronary arteries using multi-detector computed tomography (MDCT) allow for early detection of athero-sclerotic plaques and other health complication, either by means of coronary artery calcium scoring (CACS) or by means of CT coronary angiography (CTCA). These techniques are widely reported in both medical and technical papers [1], [4], [11]. Coronary artery anomalies that entail a risk of sudden death are frequently associated with artery vessels malformations. In medical practice both dynamic and static vessel's parameters can be captured during computed tomography examination.

Proposed topics are closely related to the classical biometrics domain, which is noticeably presented in the scientific investigations [10]. In our approach we propose another understanding of this term as the analysis of data from human clinical trials evaluating the relative effectiveness of therapies for disease.

2 Source Data Description

Data acquisition process starts with a non-invasive CT examination of the patient. As a result of the CT procedures raw-data in DICOM format is obtained. The raw-data is appropriately analysed and after that displayed on a graphic workstation. By means of specialized programs, the raw-data is also reconstructed and displayed as various images by a qualified radiologist. Images are built from this data and displayed on the monitors in a range of scales, in different views, three-dimensional projections and reconstructions. These projections produce the "tomogram": a two-dimensional image of a section, or slice, through a three-dimensional object, containing the cross-sections of the artery.

The cardiac CT scans present images similar to Fig. 1. The radiologist indicates the areas and sections of blood vessels where measurement should be conducted. Fig. 1 presents a CT image upon which a blood vessel examination has been performed. In this picture, different cross-sections along the vessel can be studied. Each cross-section is surrounded by a frame, which has been automatically added by the work-station software. Inside the frame, the shape of the cross-section is then estimated. This examination was performed by a Toshiba Aquilion 64 CT scanner and by the Vitrea2 (Vital Images Inc.) application at the Unit of the Non-invasive Cardiovascular Diagnostic at the Medical University of Silesia. If the reference points for the measurements have been chosen by the physician, then evaluations will be automatically conducted at these points and at points adjacent to each point, both in front of the point and behind it. Here seven locations before and seven locations after the reference point are studied: fifteen blood vessel cross-sections will be

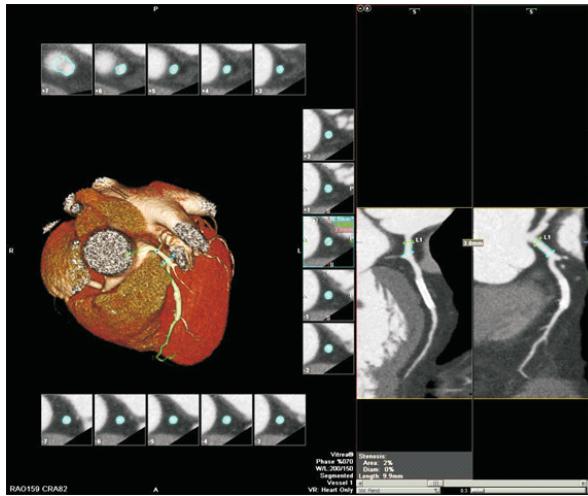


Fig. 1 The source CT image presenting 15 sub-images containing cross sections of a given vessel

investigated. The cross-sections can be separated by a constant distance; here the step size is 0.5 mm.

As was mentioned, inside every individual image I , fifteen sub-images are included: $I \supset \{I^1, I^2, \dots, I^{15}\}$; each sub-image $I^j, j = 1, \dots, 15$ includes one cross-section of the analyzed vessel [9]. For this paper, due to technical restrictions, the images and the places on these images at which measurements are conducted are too small to be properly displayed. Therefore, in Fig. 2a and Fig. 3a, the examples of magnified cross-sections have been shown. Fig. 2a presents an example of a healthy patient's cross-section. It may be noted that the shape of the cross-section resembles an ellipse. In Fig. 3a the cross-section of a patient with the pathological changes has been shown.

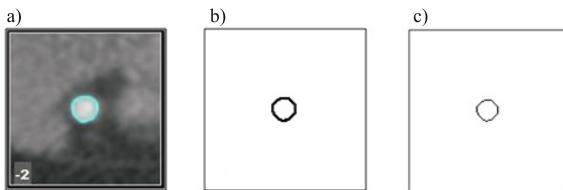


Fig. 2 Exemplary cross-section of a healthy patient's blood vessel (a), cross-section after binarization process (b) and the inner outline of the cross-section's contour (c)

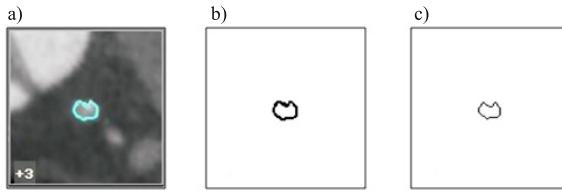


Fig. 3 An example of the cross-section with visible pathological changes (a), cross-section after binarization process (b) and the inner outline of the cross-section's contour (c)

3 Image Analysis Method

CT examination cannot display many smallest details, so in some cases it is impossible to anticipate further therapy or monitoring. Digital image produced by the CT device allows visualising and identifying the cross-sections' areas where roughness should be measured. Unfortunately it has to be measured manually what makes this analysis very time consuming. In this paper a computer aided method for measuring the static parameters (roughness) has been presented.

The purpose of data acquisition process is to obtain the CT images that can be analyzed to measure the roughness of the vessel's wall. Input images include cross-sections of an artery in the section indicated by the diagnostician, in the same scale, numbered from +7 to -7 (Fig. 1).

Contrary to the methods of measuring the vascular contractility [9] method presented in this article does not require a series of images triggered by electrocardiogram signal, and can be conducted on a single image, containing the cross-sections of the coronary artery.

Initial stages of image processing are identical to those described in the work [9]. First, the binary representation of the source CT image is established (as shown in Fig. 2b and Fig. 3b). Binarization is the stage of image processing used for various applications (such as in medicine [3], biometrics [7], [10]) in order to prepare the image for further analysis. Next, the coordinates of the four vertexes of every frame are designated. As a result of these operations the coordinates of the upper left corner and dimensions of the sides of each frame are obtained. This data allow to determine the coordinates of frames' centers which are the input data to the algorithm of filling the cross-sections.

In the subsequent step, for a given cross-sections pointed out by radiologist, their centre of gravity (x_c, y_c) is automatically determined:

$$x_c = \frac{1}{m} \sum_{j=1}^m x_j, \quad y_c = \frac{1}{m} \sum_{j=1}^m y_j \quad (1)$$

where:

x_j, y_j – current coordinate of the cross-section area's point,

m – number of points in the cross-section's area.

In the next stage the inner outline of the cross-section of the vessel wall is established. In practice, digital image of the vessel's contour consists of separated points (pixels). These pixels have to be correctly selected from the set of all points, which create vessel's cross-section. The vessel's contour is made up of pixels which form a line of non-uniform thickness (light gray and dark gray pixels in Fig. 4a). If all these pixels will be taken into account then the internal shape of the contour (dark gray pixels in Fig. 4a) will be incorrectly designated. Described disadvantage can be eliminated by means of the algorithm that creates a chain of appropriate pixels [12]. This algorithm is used to construct a chain of pixels constituting the inner outline of the vessel (light gray pixels in Fig. 4a) and is described in more detail in the next part of this work.

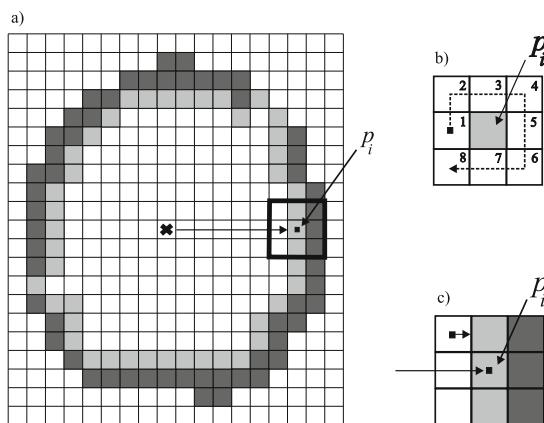


Fig. 4 The order of designating the points of vessel's inner outline (a),(c), the order of finding the next pixel in the neighbourhood of the pixel that was qualified as last (b)

An ordered sequence of pixels can be presented as a chain $P = (p_0, \dots, p_{n-1})$ of pixels. The number n is the number of the chain elements, and each pixel $p_i = (x_i, y_i)$ is determined by the following parameters:

i – current number of the pixel, $i = 0, 1, \dots, n - 1$,

x_i, y_i – coordinates of i -th pixel.

The pixels with the coordinates (x_i, y_i) and (x_{i+1}, y_{i+1}) form a sequence when the following dependence is fulfilled:

$$\forall_{p_i \in P} (|x_i - x_{i+1}| \leq 1 \wedge |y_i - y_{i+1}| \leq 1) \quad (2)$$

A sample chain of pixels is shown in Fig. 5. The process of creating pixel sequences within analyzed image is initiated by finding the first black pixel lying on the vessel's cross-section. The search begins with a pixel located in the centre of the cross section and ends when a pixel positioned to the right of the cross-section centre is found. The process of finding the first pixel is illustrated in Fig. 4a.

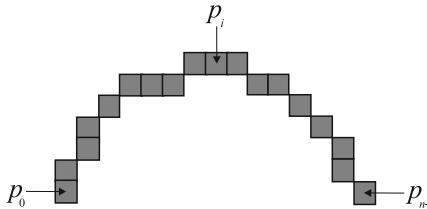


Fig. 5 Segment of pixels chain

After finding the first pixel of the sequence, its coordinates are stored for later use. The searching process of next pixels to join the sequence continues in the neighbourhood of the pixel that was qualified for the chain as last, according to the search order presented in Fig. 4b. The algorithm stops when coordinates of the two last qualified pixels are the same as the coordinates of the two first qualified pixels.

The algorithm returns the set of pixels which form the one-pixel thick inner outline of the vessel's wall (see Fig. 2c and Fig. 3c). For a given set of pixels which form a regular internal vessels' contour, the distance d_i between these points and the cross-section's centre of gravity is calculated:

$$d_i = \sqrt{(x_c - x_i)^2 + (y_c - y_i)^2} \quad (3)$$

where:

x_c, y_c – coordinates of the cross-section's centre of gravity,

x_i, y_i – coordinates of the current point of the outline, $i = 1, \dots, n$,

n – number of the pixels which form the cross-section's inner outline.

In the next stage, for all computed distances d_i , their arithmetic mean d_{avg} is calculated.

4 Roughness Estimation

The concept of roughness occurs in mechanics and materials sciences. In this paper, the principle of determining the roughness of the surface was transferred to medical image processing. Roughness is a characteristic of solid surfaces characterized by its mechanical inequality [2]. In this paper, the following methods of determining the surface roughness were used and compared:

- Standard Deviation σ :

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - d_{avg})^2} \quad (4)$$

- Roughness height according to the 10 points of the profile Rz :

$$Rz = \frac{1}{5} \sum_{k=1}^5 W_k - \frac{1}{5} \sum_{k=1}^5 D_k \quad (5)$$

where:

W_k – five maximum values (peak heights) of $(d_{avg} - d_i)$,

D_k – five minimum values (valley depths) of $(d_{avg} - d_i)$.

- Arithmetic average deviation of the profile from the average line Ra :

$$Ra = \frac{1}{l} \int_0^l |f(x)| dx \approx \frac{1}{n} \sum_{i=1}^n |d_{avg} - d_i| \quad (6)$$

where:

l – the length of the analyzed elementary section of the cross-sectional area.

All the methods of determining surface roughness have been used with the test data. The results and their discussion can be found in the next section of this work.

5 The Course and Results of the Studies

The determination of the roughness of coronary vessels' cross sections has been practically demonstrated for 149 images derived from different patients. Each image consisted of a set of 15 sub-images, as shown in Fig. 1, presenting cross sections for one patient. These images were processed by the image processing procedures described in the previous paragraphs. The parameters of the two outermost cross-sections (sub-images number +7, +6 and -7, -6) were not analyzed. Due to their close proximity to vessel branching points they may introduce serious distortions to the results. Hence, only eleven of the vessel cross-sections have been examined. The vessel cross-section roughness have been listed (see Table 1). Table 1 includes the complete results from the measurements made on one image (patient). The vessel's roughness was calculated using the methods described in the previous paragraphs. Additionally, the last row of Table 1 lists the variance of roughness values for each method separate. This measure allows to determine the dynamics of the cross-section's roughness along the whole section length of the examined vessel in the analyzed image. In case of significant changes in roughness values within a single image, the radiologist may suspect vessel calcification by deposition of atherosclerotic plaque. Observing the roughness and its variance, it can be seen that the roughness of the vessel in all tested areas is of the same order of magnitude and shows no significant differences. From the data presented in Table 1 it can be noticed that the variance of roughness calculated for the vessel has low value, especially in case of methods σ (4) and Rz (5), so, with high degree of

Table 1 Vessel cross-section roughness for a healthy patient

Cross-section No	σ	Rz	Ra
5	0,89	2,46	5,16
4	0,63	1,75	3,99
3	0,81	2,12	4,09
2	0,86	2,42	4,68
1	0,95	2,67	4,62
0	0,63	1,74	4,85
-1	0,59	1,72	4,20
-2	0,75	2,09	4,93
-3	0,71	1,98	4,67
-4	0,75	2,01	4,07
-5	0,71	1,86	3,46
Var	0,02	0,17	0,71

Table 2 Vessel cross-section roughness for patient with atherosclerotic plaque

Cross-section No	σ	Rz	Ra
5	1,97	5,94	9,84
4	2,32	6,01	7,56
3	2,53	7,72	7,53
2	1,68	4,41	9,70
1	0,71	2,50	10,04
0	0,77	2,71	9,23
-1	0,75	2,37	9,25
-2	0,81	2,58	9,35
-3	0,74	2,68	9,60
-4	0,74	2,48	8,64
-5	0,61	2,05	7,98
Var	0,41	2,92	1,89

probability, atherosclerosis disease is not present in the vessel examined. For comparison, in Table 2, data is presented from a patient that has been diagnosed as having atherosclerotic plaque. From Table 2 it can be clearly seen that for cross-sections number two, three, four and five the surface roughness calculated as σ (4) and Rz (5) has much higher values than the roughness of the other cross-sections, indicating a risk of disease. With high probability the higher values of roughness are caused by accumulation of plaque on the inner walls of the vessel in the examined areas. Also the variance is much higher than in the first case for the first method σ (4), the variance is more than twenty times higher. Location, described in Table 2 as cross-sections No 2, 3, 4 and 5 are, probably, already calcified as the vessel wall has become stiffened. If inner layer of the artery is calcified, but pathological changes are still small, then these pathologies cannot be imaged by normal CT procedures.

However, pathologies of this type can be observed by the method described in this paper. This approach can also be used for automatic cross-sectional area computation in cases when pathological changes inside vessel are already well visible on CT images.

In the case of roughness R_a (6) changes of the values are difficult to assess. This method does not seem to be adequate for the applications presented in this study.

6 Conclusions

The method proposed in this paper allows data to be automatically read from CT images generated on the basis of information received from the CT scanner. Some additional information can also be extracted from CT images. This additional information has been credibly and correctly interpreted, and the interpretation given matched the diagnosis made by a physician during the patient's examination. The presented method, unlike the method of analysis of dynamic parameters (contractibility) of coronary vessels [9], eliminates the need to prepare a series of images related to electrocardiogram, and thus to diagnose the disease after a single tomography image analysis. Due to the automatic measurements it is possible to analyze the roughness in real time. Previously, due to lack of appropriate tools, it was not possible.

Combining the methods of analysis for dynamic (contractibility) and static (roughness) parameters of the coronary arteries will create a decision support system that will provide data supporting the diagnosis of coronary disease by the physician. With a sufficiently large data set the prediction system for heart disease can be constructed [8]. The described methods can also be the basis for further studies, such as in [5] or [6]. Rapid method for obtaining additional information allows for the development of new prevention therapy methods.

The CT images dataset derive from the Medical University of Silesia archives and from current medical examinations. All these images are owned by the Medical Center. During investigations 149 CT images have been transformed and then processed, according to the methods presented in this paper. This procedure and our software is used by physicians during CT image analysis. The procedure detailed herein was accomplished with a graphical C# program tool that runs on all tomography workstations. Thus, using this new technique, CT images can be also studied at medical centres at which computed tomography is not present. This is possible as, after their examination, the patients can obtain their data and medical documentation on CD media. It should be emphasized here that this method permits the identification of pathological changes when these changes are not yet visible on CT images.

Acknowledgements. We are grateful to doc Maciej Sosnowski from Medical University of Silesia (Unit of the Noninvasive Cardiovascular Diagnostics; 3rd Division of Cardiology, Upper Silesian Heart Centre, Katowice, Poland) for his collaboration, valuable advice and providing the set of test data.

References

1. Chien, C., Feng, Y.F., Arora, R.R.: Advances in computed tomography-based evaluation of coronary arteries: a review of coronary artery imaging with multidetector spiral computed tomography. *Rev. Cardiovasc Med.* 8, 53–60 (2007)
2. Degarmo, E.P., Black, J.T., Kohser, R.A.: Materials and Processes in Manufacturing, 9th edn., p. 223. Wiley (2003)
3. Dougherty, G.: Digital Image Processing for Medical Applications. Cambridge University Press (2009)
4. Gola, A., Pysz, P., Szymanski, L., Sosnowski, M.: 64-slice MDCT coronary artery functional assessment in zero calcium score patients a novel parameter emerging. *Eur. Radiol.* 18(supl. 3) (C25), 86 (2008)
5. Koprowski, R., Wrobel, Z., Zieleznik, W.: Automatic Ultrasound Image Analysis in Hashimoto's Disease. In: Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Kittler, J. (eds.) MCPR 2010. LNCS, vol. 6256, pp. 98–106. Springer, Heidelberg (2010)
6. Koprowski, R., Wrobel, Z.: Layers recognition in tomographic eye image based on random contour analysis. In: Kurzynski, M., Woźniak, M. (eds.) Computer Recognition Systems 3. AISC, vol. 57, pp. 471–478. Springer, Heidelberg (2009)
7. Orczyk, T., Porwik, P.: Feature based face detection algorithm for digital still images. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 3. AISC, vol. 102, pp. 59–64. Springer, Heidelberg (2011)
8. Papageorgiou, E.I., Froelich, W.: Application of Evolutionary Fuzzy Cognitive Maps for Prediction of Pulmonary Infections. *IEEE Transactions on Information Technology In Biomedicine* 16, 143–149 (2012)
9. Porwik, P., Sosnowski, M., Wesolowski, T., Wrobel, K.: A Computational Assessment of a Blood Vessel's Compliance: A Procedure Based on Computed Tomography Coronary Angiography. In: Corchado, E., Kurzyński, M., Woźniak, M. (eds.) HAIS 2011, Part I. LNCS (LNAI), vol. 6678, pp. 428–435. Springer, Heidelberg (2011)
10. Porwik, P., Para, T.: Some handwritten signature parameters in biometric recognition process. In: Proceedings of the ITI 2007, 29th International Conference on Information Technology Interfaces, pp. 185–190. IEEE, Croatia (2007)
11. Schroeder, S., Kopp, A.F., Baumbach, A., et al.: Noninvasive detection and evaluation of atherosclerotic coronary plaques with multislice computed tomography. *J. Am. Coll. Cardiol.* 37, 1430–1435 (2001)
12. Wrobel, K., Doroz, R.: The method for finding a reference point in fingerprint images basing on an analysis of characteristic points. In: Third World Congress on Nature and Biologically Inspired Computing (NaBIC), pp. 504–508. IEEE Press, Salamanca (2011)

Performance Benchmarking of Different Binarization Techniques for Fingerprint-Based Biometric Authentication

Soharab Hossain Shaikh¹, Khalid Saeed², and Nabendu Chaki³

Abstract. Fingerprint analysis is a well-known biometrics for person identification and authentication. Computer vision based fingerprint recognition systems follow different preprocessing steps among which binarization is one of the most important steps. In an attempt towards benchmarking different global and local binarization techniques, it has been found that for a set of sample images two or more techniques show identical performance in terms of successful authentication. However, the accuracy of a fingerprint recognition system varies with different binarization techniques used in the preprocessing step and the same cannot be enumerated in the objective measure of success or failure. This paper proposes a quantitative evaluation measure namely confidence score to be used for benchmarking of different binarization techniques in a more effective manner.

1 Introduction

Binarization is an important step in image processing and pattern recognition applications. Binarization is also an important image segmentation technique as it reduces the dimension of the image resulting in less computational overhead and lower memory requirement compared to processing the original color or gray-scale image. The recognition accuracy of a system highly depends on the performance of

Soharab Hossain Shaikh

A.K.Choudhury School of Information Technology, University of Calcutta, Kolkata, India
e-mail: soharab.hossain@gmail.com

Khalid Saeed

Faculty of Physics and Applied Computer Science,
AGH University of Science and Technology, Cracow, Poland
e-mail: saeed@agh.edu.pl

Nabendu Chaki

Department of Computer Science and Engineering, University of Calcutta, Kolkata, India
e-mail: nabendu@ieee.org

underlying image segmentation method, like binarization. In the view of this, the present paper reports the finding of the authors on the performance benchmarking of different binarization techniques for fingerprint recognition systems. A number of metrics are available in the literature [9] [10] [11] like false rejection rate (FRR), false acceptance rate (FAR), equal error rate (EER) for analyzing the performance of a fingerprint recognition system. In this paper, the authors propose a new coefficient of measure for evaluating six different binarization methods in the context of the biometric application - fingerprint recognition. In the proposed approach the numeric estimation of performance for the new measure is done only after observing the impact of the binarization technique on producing the final objective of correct fingerprint recognition. Two global binarization methods namely Otsu [3] and Kapur [4] and four local methods - Bernsen [5], Niblack [6], Sauvola [7], and Iterative partitioning [1] have been used for the experimentation purpose. In the current scenario, it is often found that even if an algorithm identifies the correct solution, the second best result for a wrong sample is actually quite close. On the other side, often an algorithm fails to identify the correct solution by a very narrow margin. Thus, a numeric count of success percentage does not quite reflect how good or bad had been the algorithm. While using some binarization method leads to the identification of the correct sample, usage of some other method may lead to a situation where one or more rejected sample is found to be in close proximity with the correct sample. The present paper addresses this concern and proposes a new coefficient of measure for evaluation of the goodness of a binarization algorithm in the context of a fingerprint-based biometric authentication application. This coefficient reflects the level of confidence of a binarization algorithm in identifying the correct sample. The objective of this paper is not to check for the superiority of a particular binarization algorithm to others. Rather, this paper aims at finding the binarization method that best suits a typical application (fingerprint recognition at the present work) based on a new coefficient of measure, confidence score.

2 Previous Works

This section briefly presents a very short introduction of binarization techniques and fingerprint evaluation metrics.

2.1 *Binarization Techniques*

Binarization is the process of converting an image from color or gray-scale to bi-level representation. In a binary image only two bits are used for representing the image. Generally an object is represented with zero and the background with one or vice versa. There are basically two types of techniques for binarization: global and local depending on how the threshold value is selected from grouping the pixel intensities into two groups (foreground and background). In the present paper the authors have considered two global binarization methods (Otsu and Kapur) and four

local binarization methods (Bernsen, Niblack, Sauvola and Iterative Partitioning). More detailed description of binarization can be found in [1] [2]. An innovative method for finger vein image binarization is presented in [8]. The paper presents an innovative approach to solve the problem of vein extraction from the inside part of hand fingers. The algorithm is easy to implement. However, the process in [8] involves multiple binarization phases using static pre-specified threshold value.

2.2 *Fingerprint Evaluation Metrics*

Human authentication using fingerprints is a well-established biometrics as found in the literature [10] [13] [14] [15] [16]. In most of the fingerprint recognition systems, performance evaluation is done using two well known metrics namely False Rejection Rate (FRR) and False Acceptance Rate (FAR). For a fingerprint image database, each sample is matched against the remaining samples of the same finger to compute the False Rejection Rate. All the scores for such matches are composed into a series of correct scores. Also the first sample of each finger in the database is matched against the first sample of the remaining fingers to compute the False Acceptance Rate. All the scores from such matches are composed into a series of incorrect scores. At some point in time, these two error rates become identical (equivalent), the error rate at that point is referred to as the Equal Error Rate (EER). These errors mainly occur in the verification application context and so, they are referred to as verification errors. In the context of an identification application, two errors can occur: (i) the matching module can produce an empty candidate list even when there is at least a template of finger whose feature set is being matched. This is known as the false negative identification error. (ii) is the false positive identification error, where the matching module returns a non-empty candidate list even when there is no single template of a finger whose feature set is being matched. False-Negative Identification-error Rate (FNIR) and False Positive Identification-error Rate (FPIR) are the commonly used terms with respect to the above mentioned errors. FNIR is the probability of a false negative identification while FPIR is the probability of the false positive identification error. Different metrics for fingerprint evaluation can be found in [9] [10] [11] [12] [13].

3 Methodology for the Proposed Work

This section briefly presents the techniques used by the fingerprint recognition system. The approach presented in [11] has been adopted in this text. However, the pre-processing step of sample image binarization in [11] has been replaced by six different binarization methods one after another towards evaluating their comparative performance in terms of the proposed new coefficient of measure, confidence score.

3.1 Fingerprint Recognition System

Minutia-based fingerprint recognition [15] represents the fingerprint by its local features like terminations and bifurcations. This approach is the backbone of many available fingerprint recognition products. There are two basic steps of minutia based fingerprint recognition systems - minutia extraction and minutia matching.

3.1.1 Minutia Extraction

In this step a series of tasks have been performed namely preprocessing, minutia extraction and post-processing. The preprocessing step consists of image enhancement followed by binarization and subsequent image segmentation. The sample images are enhanced by histogram equalization and subsequently Fourier transform is done. In the original work [11], an adaptive binarization technique [12] has been used. In the present work the authors adhere to the adaptive binarization technique in the training phase. This subtask (binarization) has been replaced by different well-known methods (Otsu, Kapur, Bernsen, Niblack, Sauvola and Iterative partitioning) for evaluating their performance during testing phase. During image segmentation, block direction estimation is done with a block size of 16x16 and extraction of region of interest using morphological operations. Minutia extraction consists of thinning of the preprocessed samples and subsequent minutia marking. The post-processing is useful for removing the detection of false minutia.

3.1.2 Minutia Matching

Given two sets of minutia of two fingerprint images, the minutia match algorithm [11] [16] determines whether the two minutia sets are from the same finger or not. This step consists of two consecutive stages: the alignment stage and the match stage [11]. In case of any two minutia from different images, if they are in a box with small length and their direction has large consistence, they are matched minutia. Match score is calculated by finding the number of matched minutia and dividing it by the maximum of number of minutia of the two images [11].

4 Proposed Evaluation Method

The proposed coefficient of measure is named as confidence score. The confidence score is measured based on a pre-computed numeric score which is another measure depending on the particular application being considered. As for example, in this paper a minutia-based matching score has been calculated for the fingerprint matching (described in section 3.1.2). This score is further used to compute the confidence score. The confidence score for successful match is formulated as shown in equation 1 and that for the unsuccessful match is presented in equation 2. In equations

1 and 2, the best score is the largest value of the matching score obtained when a test sample is compared with three different impressions of the same finger from the training set. The second best score is the second largest value of the matching score. In case of a successful match, the best score and the correct score are the same and this is the value that identifies the correct sample. The confidence score is defined with equation 1 and equation 2 for the cases of successful (α_{suc}) and unsuccessful (α_{fail}) recognitions respectively. The best score, second best score and correct score are represented by β , ζ and τ respectively, while α stands for the confidence score.

$$\alpha_{suc} = \frac{\beta - \zeta}{\beta} \times 100\% \quad (1)$$

$$\alpha_{fail} = \frac{\tau - \beta}{\tau} \times 100\% \quad (2)$$

The second best score ζ obviously correspond to the closest wrong sample. Thus the confidence score would be a non-negative value signifying the distance between the best and the second best cases. However, in case of a recognition failure, the best score β corresponds to a wrong sample. The correct score τ on the other hand is the value found against the correct sample. Obviously, in this case the correct score τ cannot be greater than the best score β . Thus the confidence score α_{fail} would be non-positive for failure while the confidence score α_{suc} would be positive for the successful authentications. The proposed confidence score finds out if a test-sample is correctly identified, how convincing the underlying classifier is in finding the result. In addition to that, the confidence score also reflects that if any testing sample is incorrectly identified, how far the matching score does lie from the second best score. A close proximity of the best score and the second best scores reflects that a small change in the input image may significantly affect the final result by correctly (or incorrectly) classifying the test-sample. On the other hand if the best and the second best scores vary significantly, it suggests that the classifier is performing confidently and a subtle change in the input test image will not affect the end result of classification. In the context of the present paper, the authors used this coefficient of measure for evaluating the performances of six binarization algorithms in the view of a minutia-based fingerprint recognition system. There are a number of processing steps of the fingerprint recognition system as explained in section 3. Among all these steps, during experimentation, the binarization step has been changed with six different binarization algorithms one-by-one. Thus, the change in the confidence score is due to the change incorporated by the results of the binarization algorithms. Hence, the confidence scores indirectly measures the performances of the binarization algorithms. The matching score used in computing the confidence score is a numeric measure which is application dependent. In the present work, the authors have used a minutia-based matching score for this purpose. This metric is described in section 3.1.2.



Fig. 1 Sample fingerprint images from the experimental dataset

5 Experimental Verification

The system consists of two phases - training and testing. In the training phase a database has been created. The experiments have been done on FVC2000 dataset [17]. Finger impression of each finger is collected for three times.

An adaptive thresholding method is used during training phase. The image is logically partitioned into 16×16 blocks and that particular block has been binarized using mean of the intensities of the pixels in the 16×16 block. In the testing phase, the binarization step has been replaced by the candidate six binarization techniques, Otsu [3], Kapur [4], Bernsen [5], Niblack [6], Sauvola [7] and Iterative partitioning [1]. The fingerprint identification process is continued with the output from each of these methods separately for each sample. The fingerprint features are extracted from the samples based on minutia based approach. A matching scores are calculated comparing the training samples with the testing samples.

5.1 Experimental Results

Fingerprint samples from FVC2000 dataset have been used for experimental work in the present paper. Three different images for each sample have been taken. The experimental results are tabulated in Tables 1 and 2. Figure 2 shows a graphical representation of the average performance of different binarization techniques.

Table 1 shows partial results when Iterative partitioning method is used for the binarization purpose. There are three impressions of each finger in the training set. During testing, when the third impression of sample 102 is given (this sample is preprocessed by Iterative partitioning algorithm) it gives a value of 61.11 as the matching score (this is the best score obtained; it is also the correct score as this is a correct classification). The second best score is 19.51. Consequently the corresponding value of the confidence-score is 68.07. This large positive value denotes that the best matching score is far from the second best one signifying that a subtle change in the input image (due to noise incorporated during sample collection or other noises like slight illumination change) less likely to affect the final classification result. This ensures better noise immunity of the underlying technology (here the effectiveness of the binarization algorithm). A large value (double digit positive number) of confidence score is also obtained for sample 109 as shown in

Table 1 Performance Evaluation of Iterative Partitioning (partial results)

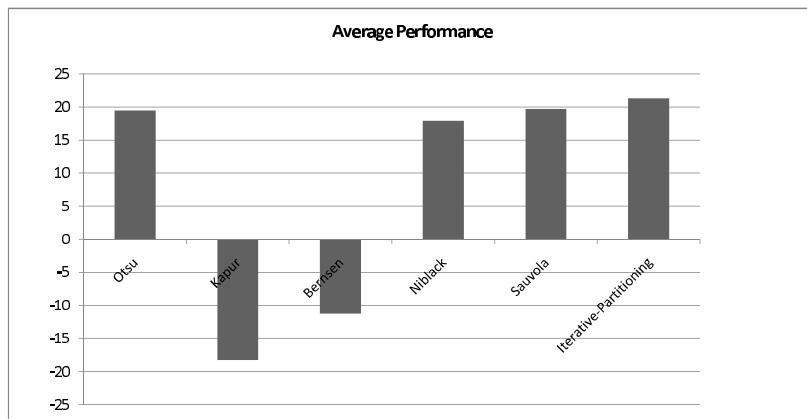
Training Sample	Impression No.	Test Sample	Impression No.	Matching Score	Confidence Score
102	1	102	3	19.51	68.07
	2			19.30	
	3			61.11	
109	1	109	1	69.67	11.67
	2			53.13	
	3			61.54	
106	1	106	3	15.00	-5.56
	2			16.67	
	3			15.79	
107	1	107	1	38.10	-3.13
	2			39.29	
	3			37.50	

Table 1. It is worth mentioning that for sample 106 the best and the second best scores are 16.67 and 15.79 respectively which are very close to each other. The negative confidence score signifies a misclassification (the test sample is actually the third impression which is incorrectly classified as the second impression by the largest matching score). The close proximity of best and the second best matching scores denotes less noise immunity. A small negative value of the confidence score suggests that the system (here the binarization algorithm) fails to detect the correct sample by a very narrow margin which is also shown for sample 107.

Table 2 summarises the results of computing the confidence score using selected fifteen candidate test samples from the FVC2000 dataset [17]. The numeric value represents the value for the confidence score obtained when a test sample is compared with a training sample. The values in the last row are the average scores. Larger value of the metric represents better score. As shown by the results, the Iterative partitioning algorithm outperforms the other binarization methods. Method proposed by Sauvola gives the second best result. The average performance of Otsu's method is very close to that of Sauvola's method. The Niblack's method stands fourth. The other two methods proposed by Bernsen and by Kapur fail miserably giving a negative confidence-score. The average confidence-score became negative for these two methods for the reason that in most of the cases, these two methods incorrectly identified the wrong sample as the correct one, giving negative values for the confidence score which collectively contributed in the negative average score. For the experimental dataset presented in Table 2, Otsu's method fails to correctly classify for three samples. Kapur's method and Bernsen's method fail for ten and eleven samples respectively. While the number of misclassifications for Niblack's method, Sauvola's method and that of Iterative partitioning is three. Figure 2 shows a plot of the performance of the six binarization algorithms based on the average confidence scores.

Table 2 Performance Evaluation of Different Binarization Techniques

	Otsu	Kapur	Bernsen	Niblack	Sauvola	Iterative Partitioning
Image1	42.86	-50	40	55	59.18	35.92
Image2	56.22	-11.89	-12.2	11.42	31.5	30.68
Image3	18.5	-13.1	-10.83	5.46	11.18	18.5
Image4	68.42	-47.37	-2.44	-5.26	40.35	68.07
Image5	20.3	7.69	28.57	23.79	25.48	29.07
Image6	17.78	-10.12	-2.11	10.14	16.69	17.78
Image7	30.29	-47.62	-24.55	29.4	32.9	30.29
Image8	15.32	25	-12.5	8.06	20.97	15.32
Image9	28.57	-21.21	-14.29	38.1	43.55	28.57
Image10	25.93	55.56	38.27	44.44	16.67	25.93
Image11	-5.56	-42.5	5	5.26	-5.56	-5.56
Image12	-3.13	27.08	-118.61	34.38	-5	-3.13
Image13	5	-128.57	-42.86	-23.08	5.41	18.66
Image14	-36.48	-54.69	-7.14	-33.33	-28.91	-1.85
Image15	7.69	37.69	-33.28	65.38	30.77	11.67
Average	19.45	-18.27	-11.26	17.94	19.68	21.33

**Fig. 2** Evaluation of Average Performances of Different Binarization Techniques

In the present paper, the proposed coefficient of measure is used for finding the desirability of a binarization method for biometric fingerprint recognition. The method (Iterative partitioning binarization method) producing the largest positive average confidence score is found to be the best candidate for the purpose.

In general, the proposed coefficient of measure can be used for deciding which among a number of alternative techniques work best for a particular system. A numeric measure for evaluating the performance of the concerned system is required.

The confidence scores can be computed using equation 1 and equation 2. The system can be tested with different techniques. The most suitable technique is the one which gives the largest value of the confidence score for the average performance. A large positive average confidence score is desirable.

6 Conclusion

This paper is primarily aimed to benchmark different binarization techniques to be used as a pre-processing step in a fingerprint recognition system. In the process, a new coefficient of measure has been proposed for quantitative evaluation of the existing binarization techniques by observing the end-result of person authentication. This new coefficient of measure is being termed as confidence score. This paper proposes the formula to compute the confidence score and has demonstrated its impact on several existing binarization techniques e.g. Otsu, Kapur, Niblack, Bernsen, Sauvola and Iterative-partitioning methods. The relative performances have been studied, documented and the experimental results are illustrated. However, the main emphasis of the paper is to establish the effectiveness of the proposed new coefficient of measure. Further experimental evidences for performance benchmarking of binarization techniques using the proposed coefficient on a larger sample size would supplement the primary objective of this paper.

Acknowledgements. The authors kindly acknowledge Mr. Wu Zhili for the code for computing the matching scores.

References

1. Shaikh, H.S., Maiti, K.A., Chaki, N.: A new image binarization method using iterative partitioning. Springer Journal on Machine Vision and Applications(MVAP) 24(2), 337–350 (2013)
2. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. J. Electr. Imaging 13(1), 146–165 (2004)
3. Otsu, N.: A threshold selection method from gray-level histogram. IEEE Transaction on Systems Man Cybernetics 9, 62–66 (1979)
4. Kapur, N.J., Sahoo, K.P., Wong, C.K.A.: A new method for gray-level picture thresholding using the entropy of the histogram. CVGIP 29(3), 273–285 (1985)
5. Bernsen, J.: Dynamic thresholding of gray level images. In: Proceedings of the International Conference on Pattern Recognition (ICPR), pp. 1251–1255 (1986)
6. Niblack, W.: An Introduction to Digital Image Processing, pp. 115–116. Prentice Hall, Eaglewood Cliffs (1986)
7. Sauvola, J., Pietikainen, M.: Adaptive document image binarization. Pattern Recognition 33(2), 225–236 (2000)

8. Waluś, M., Kosmala, J., Saeed, K.: Finger Vein Pattern Extraction Algorithm. In: Corchado, E., Kurzyński, M., Woźniak, M. (eds.) HAIS 2011, Part I. LNCS, vol. 6678, pp. 404–411. Springer, Heidelberg (2011)
9. Kaggwa, F.: Evaluation of Multiple Enrollment for Fingerprint Recognition, M.Sc. Thesis, University of Twente (2011)
10. Maltoni, D., Maio, D., Jain, K.A., Prabhakar, S.: Springer professional computing. In: Handbook of Fingerprint Recognition. Springer (2009)
11. Zhili, W.: Fingerprint Recognition, Thesis for BS in Computer Science, Hong Kong Baptist University (2002)
12. Jain, C.L., Halici, U., Hayashi, I., Lee, B.S., Tsutsui, S.: Intelligent biometric techniques in fingerprint and face recognition. CRC Press (1999)
13. Hong, L., Wan, Y., Jain, K.A.: Fingerprint Image Enhancement: Algorithms and Performance Evaluation. *IEEE Transactions on PAMI* 20(8), 777–789 (1998)
14. Jain, K.A., Hong, L., Pankanti, S., Bolle, R.: An identity authentication system using fingerprints. *Proceedings of the IEEE* 85(9), 1365–1388 (1997)
15. Maio, D., Maltoni, D.: Direct Gray-scale Minutiae Detection in Fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(1), 27–40 (1997)
16. Hong, L.: Automatic Personal Identification Using Fingerprints, Ph.D. Thesis (1998)
17. FVC2000: Fingerprint Verification Competition,
<http://www.csr.unibo.it/research/biolab>(accessed in January 2013)

Biometric Features Selection with k -Nearest Neighbours Technique and Hotelling Adaptation Method

Piotr Porwik and Rafal Doroz

Abstract. The presented work focuses on the method of handwritten signature recognition. For those objects recognition process can be difficult because repeatability of signature features is low. The most important advantage of the proposed solution is individual adjustment of features and analysis methods which distinguish each single signature inside the dataset of signatures' features. In presented approach different features and similarity measures can be freely selected. Additionally, selected features and similarity measures can be different for every person.

1 Introduction

The specification of human signatures is one of the greatest problems in the designing the credibly classifiers which work in the biometric identification or verification systems [1], [2], [6], [11]. Repeatability of signatures even the same person characterizes a large discrepancy. Some additional difficulties can be observed when someone is trying to imitate signature and steal own identity. Signature belongs to behavioral biometrics and modernly is widely acceptable and collectable characteristic [3]. It should be noticed that another different knowledge acquisition and representation techniques can be also applied [12], [13].

The algorithm, presented in this work, not only performs the selection of the signature features but also indicates the best similarity measures (from the set of all available measures). It allows minimizing signature verification error. The most important advantage of the proposed approach is possibility of a choice different signature features and different similarity measures for every individual. In the proposed method of signature classification, the two stages can be distinguished: training and verification mode. The aim of a training stage is to create training sets. Thanks to

Piotr Porwik · Rafal Doroz

Institute of Computer Science, University of Silesia, ul. Bedzinska 39, 41-200 Sosnowiec
e-mail: {piotr.porwik, rafal.doroz}@us.edu.pl

these sets it is possible to evaluating which features and methods of their analysis the best distinguish the original signature of a given person from the forged signatures. The best measures, distinguishing signature from others are then collected and connected with a given person. This information will be then used in the classification attempts.

2 The Hotelling's Statistic

The accuracy of the system depends mainly on the quality and number of the object's features. For this reason only selected number of features should be retrieved, then recognition level of objects will be acceptable. Additionally, majority of recognition algorithms have data dimensionality limitation, hence only restricted data can be practically processed. In this paper that idea is also applied [4], [6].

In proposed in this paper solution the dimensionality data reduction was performed by means of the Hotelling's T^2 statistic. In many statistical methods, the possibility of using a particular method is conditioned by the fact that the data must to have specified probability distribution. For this reason, the nature of the distribution is one of the major tasks in biometrics. In the fundamental definition of the Hotelling's statistic we have n independent vectors of dimension p [7]. The main goal of the paper is to analyze and recognition of two kinds of signatures (objects) - the original and forged signatures. These sets of signatures form two classes of objects. In the two-sample problem we have the two sets of independent vectors of features which form the two observation matrices:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}^T \quad \mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{p1} & y_{p2} & \cdots & y_{pm} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_m \end{bmatrix}^T \quad (1)$$

The features create the vectors \mathbf{x}_i , $i = 1, \dots, n$ and \mathbf{y}_j , $j = 1, \dots, m$ and form a p -dimensional normally distributed population $\mathbf{x}_i \sim N_p(\mu', \Sigma)$, $\mathbf{y}_i \sim N_p(\mu'', \Sigma)$. The parameters distribution are unknown and have to be estimated. The parameter μ can be estimated by a mean vectors:

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad \text{and} \quad \bar{\mathbf{y}} = \frac{1}{m} \sum_{i=1}^m \mathbf{y}_i. \quad (2)$$

The parameter Σ for the two-sample multivariate case has to be formed as a maximum likelihood estimator [7]:

$$S = \frac{1}{n+m-2} \left(\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T + \sum_{i=1}^m (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \right). \quad (3)$$

The vector of the mean values in the all observation matrices is defined as follows:

$$\bar{z} = \frac{1}{n+m} \left(\sum_{i=1}^n \mathbf{x}_i + \sum_{i=1}^m \mathbf{y}_i \right). \quad (4)$$

For such assumption a two-sample T^2 Hotelling statistic can be defined as follows:

$$T^2 = \frac{1}{n+m-2} \left(n \left[(\bar{\mathbf{x}} - \bar{\mathbf{z}})^T S^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{z}}) \right] + m \left[(\bar{\mathbf{y}} - \bar{\mathbf{z}})^T S^{-1} (\bar{\mathbf{y}} - \bar{\mathbf{z}}) \right] \right). \quad (5)$$

The two-sample Hotelling's T^2 statistic can be approximated by means of the F -Snedecor's distribution [7]:

$$\tilde{F} = \frac{n+m-p-1}{p(n+m-2)} T^2, \quad (6)$$

then $\mu = 0$ if for significance level α , the next condition is fulfilled $\tilde{F} > F_{p,n+m-p-1,\alpha}$.

3 Signature Features Preparation

In the first step, the two sets of signatures are created for every person. Let a set containing only original signatures be denoted as GS and forged signatures of the same person be denoted as GS^* , then:

$$GS = \{S_1, S_2, \dots, S_c\} \quad \text{and} \quad GS^* = \{S_1^*, S_2^*, \dots, S_d^*\}. \quad (7)$$

The forged signatures are always difficult in practical acquisition and are accessible in particular cases only – for example during police investigations. In proposed approach this kind of signatures will be randomly generated. As aforementioned, the signature verification can be carried out on the basis of different signature features and different similarity measures. The set of all available signatures' features will be denoted as follows:

$$F = \{f_1, f_2, \dots, f_u\}. \quad (8)$$

In other words, the set F contains the global number of signature features which can be measured and captured during acquisition process. The features can be even substantially different for every signature. Signatures have to be recognized by different methods, factors or similarity coefficients. Let a set of these methods be denoted as:

$$M = \{\omega_1, \omega_2, \dots, \omega_k\}. \quad (9)$$

It means that the set M contains all known methods and mathematical rules which can be included to the classification process. Each signature's feature can be recognized by means of any acceptable method but only one (or a few) of them should

give the best discriminant factor. These discriminant factors are not known *a priori* and have to be discovered in future. Signature data preparations have been deeply presented in the work [1].

Let for a single signature, a set of all possible combinations of pairs “feature-method” be defined as follows:

$$FM = \{\varepsilon_i = (f_m, \omega_j)_i : f \in F, \omega \in M\}, i = 1, \dots, u \cdot k, \quad (10)$$

where:

$(f_m, \omega_j)_i$ - the i^{th} pair “signature’s feature (f_m) - analysis method (ω_j)”, $m = 1, \dots, u$, $j = 1, \dots, k$,

u - number of the signature features,

k - number of methods used in comparison of the features.

In this paper selected parameters were analyzed and it presents Table 1.

Table 1 List of analyzed features

Description of the feature f_i	Name of the signature feature recorded by pen (table) i
1	pressure
2	acceleration
3	x – velocity
4	y – velocity
5	x,y – velocity
6	x – coordinate
7	y – coordinate

Table 2 presents various similarity measures (coefficients) - each of them can have different range of variability. Measurement difficulties were overcome in the normalization process, what reports the work [7]. To accomplish this task the well known Dynamic Time Warping (DTW) technique was applied [9].

4 Data Sets Preparation

Based on the learning sets, classifier performs verification of a new, unknown signature and recognizes this signature from a database. In the proposed method, the learning sets consist of the two matrices \mathbf{X} and \mathbf{Y} . Thanks to them, the classifier is able to distinguish the original signatures of a given person from other forged signatures. The correct creation of the training set (the set of examples) is very important and significantly affects the classifier effectiveness. The matrix \mathbf{X} is built on the basis of the set GS of original signatures of the person. The matrix contains the values

Table 2 Short list of similarity measures or coefficients [8]

Description of the measure ω_i	Name of coefficient or similarity measure
i	
1	Euclidean
2	Gower
3	Minkowski
4	SÄyrensen
5	City Block
6	Soergel
7	Kulczynski

of the similarity coefficient Sim calculated between pairs of original signatures. As the similarity coefficient Sim any, normalized to the interval $[0,1]$, symmetric similarity measure can be taken. Generally, the matrix \mathbf{X} has the following structure:

$$\mathbf{X} = [[S_1 \leftrightarrow S_2], \dots, [S_1 \leftrightarrow S_c], \dots, [S_{c-1} \leftrightarrow S_c]], \quad (11)$$

where:

S_i, S_j - the i^{th} and j^{th} th original signature of a given person,

c - number of all genuine signatures of a given person.

The second matrix \mathbf{Y} is created on the basis of the set GS^* . This collection includes different types of forged signatures only of the person Q . In practice, the set GS^* consists of originally forged signatures (if such are accessible) or/and includes the randomly selected signatures other persons from the set GS . The matrix \mathbf{Y} is built as follows:

$$\mathbf{Y} = [[S_1 \leftrightarrow S_1^*], \dots, [S_1 \leftrightarrow S_d^*], \dots, [S_c \leftrightarrow S_d^*]], \quad (12)$$

where:

S_i, S_j^* - the i^{th} genuine and the j^{th} forged signature of a given person,

d - number of all unauthorized (falsified) signatures of a given person.

5 Hotelling's Features Reduction Method

We assumed that analyzed specific data have the normal distribution, what, for signatures, follows from the investigations reported in the work [7], [11]. In practice, discriminant analysis is useful to decide, whether selected pair “feature-method” is important for the classification process. In this procedure, from the all possible pairs only pairs with greatest discriminant power will be left. As was mentioned above the set FM contains all possible pairs “feature-method” which can be applied during recognition process. Some recognition methods included in the set “feature-method”

can have better discriminant properties than other. It means that some signature's features can be better recognized by means of specific methods or some features cannot be classified - then should be rejected. These properties are not known *a priori*, and have to be discovered in the future analysis. For a given signature only its best discriminant features and methods of their recognition will be ultimately selected. Reduction features can be carried out gradually. In each successive step is eliminated such feature that causes the smallest reduction of a multidimensional measure of discrimination. The algorithm of reduction of the matrices dimension can be executed in several steps. In practice, it leads to the data reduction in the matrices \mathbf{X} and \mathbf{Y} . Reduction of the dimension of the matrices \mathbf{X} and \mathbf{Y} were performed step by step for every accessible pair "feature-method".

1. Taking into consideration absence of the i^{th} pair "feature-method", the discriminant measure is computed:

$$\forall i \in \{1, \dots, \#FM,\} T_i^2(\varepsilon_1, \dots, \varepsilon_{i-1}, \varepsilon_{i+1}, \dots, \varepsilon_p), \quad (13)$$

where p , as previously, denotes number of all possible combinations of the pairs "feature-method". In the first step of the algorithm $p = u \cdot k$ is assumed. Hence, the matrices \mathbf{X} and \mathbf{Y} reduction procedure concerns always reduction of the current pair (f, ω) .

2. The necessity of the i^{th} pair "feature-method" can be checked as follows:

$$U_i = T^2(\varepsilon_1, \dots, \varepsilon_p) - T^2(\varepsilon_1, \dots, \varepsilon_{i-1}, \varepsilon_{i+1}, \dots, \varepsilon_p), \quad (14)$$

where the difference between two Hotelling statistics are calculated.

3. The value of the F -test is calculated:

$$\tilde{F} = (n + m - p - 1) \cdot \frac{U_i}{1 + T^2(\varepsilon_1, \dots, \varepsilon_p) - U_i}. \quad (15)$$

4. The value \tilde{F} from (15) is then compared with the table of critical values of the Snedecor's F -distribution. In our approach, in the each loop of the algorithm the significance level $\alpha = 0.05$ was established. Because we have two classes only, then the i^{th} pair is redundant (can be rejected) if the following condition will be fulfilled: $\tilde{F} < F_{1, n+m-p-1, \alpha}$.

If the value of F does not fall into the critical region, the current i^{th} row of the matrices \mathbf{X} and \mathbf{Y} can be removed. In the successive step the parameter p is decreased, and algorithm always starts with the new value of the parameter p from the beginning, that is for $p - 1, p - 2, \dots$. Hence matrices dimension is also successively reduced. Finally, remaining elements of the matrices \mathbf{X} and \mathbf{Y} have a biggest influence on the classifier's work. The mentioned procedure is executed for the every person Q . Computation results are stored in the set $FM^{(Q)}$. It means that for the person Q we obtain the set $FM^{(Q)} \subset FM$ in which only selected pairs $(f, m)^Q$ are included. These pairs the best distinguish the genuine signature of the person Q from his/her forged signatures.

6 Verification Procedure

In the proposed approach well known the k -Nearest Neighbour (k -NN) method was applied [5]. Genuine signatures came from the database, so these signatures forms the first class (π_1) of objects. The falsifications form the second class (π_2) of the signatures. If original forged signatures are not available, genuine signatures of the other users can be used as unauthorized. These classes have to be correctly recognized during classification process. Determination of the training sets $FM^{(Q)}$ is a necessary condition for verification of the signatures taken by the person Q . Let S^Q be a signature of a person Q which should be verified. This person appears himself as a person Q . It should be automatically verified. Let the reduced matrixes \mathbf{X} and \mathbf{Y} be denoted as $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ have the same number of rows (say r) and $r \ll p$ but can have different number of columns. Hence, a new matrix $\mathbf{H} = [\tilde{\mathbf{X}} \tilde{\mathbf{Y}}]$ is formed. The matrix \mathbf{H} includes similarities between all signatures (say person Q) stored in the database, but these similarities have been computed on the base of selected features and selected recognition methods only. Columns of the matrix \mathbf{H} can separately be treated as the vectors \mathbf{h} , then:

$$\mathbf{H} = [\mathbf{h}^1, \mathbf{h}^2, \dots, \mathbf{h}^l], \quad \mathbf{h}^i = [h_1^i, h_2^i, \dots, h_r^i]^T. \quad (16)$$

If $\mathbf{h}^j \in \tilde{\mathbf{X}} \rightarrow \mathbf{h}^j \in \pi_1$ and if $\mathbf{h}^j \in \tilde{\mathbf{Y}} \rightarrow \mathbf{h}^j \in \pi_2$, $j = 1, \dots, l$.

After reduction, the obtained vectors \mathbf{h} include similarities between signatures of the same person. Because the classifier works in the verification mode, the classified person Q appears himself as a person Q . Bearing in mind this event, from the database, the signature S^Q of the person Q is randomly selected. On this stage of verification, the common most distinctive features and signature similarity measures of the person Q are just established; it means that reduced matrix \mathbf{H} for this signature is known. For this reason a new vector $\mathbf{h}^Q = [h_1^Q, h_2^Q, \dots, h_r^Q]^T$ is only created.

In the next stage, a set of all the Euclidean distances $d(\mathbf{h}^Q, \mathbf{h}^j)$ to the all signatures of the person Q is successively calculated:

$$\left\{ d(\mathbf{h}^Q, \mathbf{h}^j) \right\} = \left[\sum_{i=1}^r (h_i - h_i^j)^2 \right]^{1/2}, \quad j = 1, \dots, l. \quad (17)$$

Among all distances the smallest k distances are selected [5]. Then k -NN classifier is applied. Selected distances may belong to the two classes: π_1 (original signature) and π_2 (forged signature). The verified signature S^Q is classified to the class π_1 or π_2 by using the simple voting method. Let $d1$ and $d2$ be the set of number which a show how many times signature S^Q was classified to the class represents π_1 and π_2 :

$$d1 = \{d(\mathbf{h}^Q, \mathbf{h}^j \in \pi_1)\}, \quad d2 = \{d(\mathbf{h}^Q, \mathbf{h}^j \in \pi_2)\}, \quad (18)$$

then classification voting principle can be formulated as follows:

$$S^Q : \begin{cases} \text{genuine signature of the person } Q & \text{for } \#d1 > \#d2 \\ \text{forged signature of the person } Q & \text{for } \#d1 \leq \#d2 \end{cases}. \quad (19)$$

Signatures came from the SVC2004 database [14]. It contains both genuine signatures and skilled forgery signatures. In the experiments the signatures features and similarity coefficients from Table 1 and Table 2 were taken. The researches were carried out with the use of a database containing 800 signatures coming from 40 people. For each person, its 10 original signatures and 10 professionally forged were used. After reduction process, similarities between original and forged signatures have been judged by k -NN classifier. The final classification results are established on the basis of voting score - it depends on the number of the k neighbors which belong to a class π_1 or π_2 . The details of the method will be explained by example, where only selected signature's features and similarity measures were applied. This restriction follows from the reader's perceptual possibilities only. Fig. 1a plots the similarity distribution between different signatures of the person Q . In practice, each point (triangle or circle) in this plot has individual label. For example label (1-5) represents similarity between original signatures no.1 and no.5 of the person Q , while label (8-5F) represents similarity between original signature no.8 and forged signature no.5 of this person. For simplification only selected labels have been shown. Fig.1a presents also decision-making areas for the trained classifier and show to which class would be assigned a point with specific coordinates. Signature of the person Q will be recognized as an individual, who want to obtain an unauthorized access to the resources. Fig.1b depicts k -NN classifier decision area during classification of the individual Q . Signature of the person Q will be recognized as an individual, who want to obtain an unauthorized access to the resources.

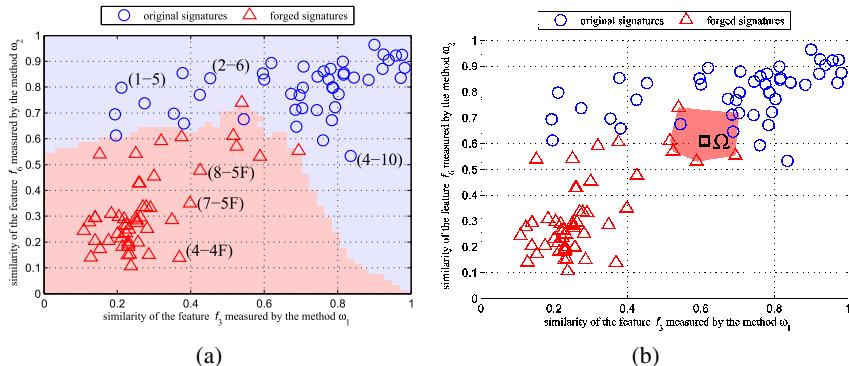


Fig. 1 Mutual similarities between signatures of the person Q (a); k -NN verification area of the person Q (b)

From Fig. 1 clearly follows that all original and forged signatures of the individual Q are well separated and can be correctly recognized during classification process.

7 Results Obtained

Results obtained were also compared with other reported solutions, where various signatures recognition techniques were applied. These achievements present, in a brief overview, Table 3. Unfortunately, reliable comparison of various approaches is quite difficult, which is caused by inconsistency in presented standards. For this reason results announced are not unified. Signatures can be captured in *on-line* or *off-line* scheme [4], [6], [10]. For this reason both capturing methods were compared.

Table 3 Comparison of performance of various signature recognition systems [10]

The approach	FAR	FRR	Signature recognition system	
			<i>off-line</i>	<i>on-line</i>
Proposed approach âĂŞ Hotelling's reduction and k-NN	1.08	2.53	+	+
Exterior Contours and Shape Features	6.90	6.50	+	
HMM and Graphometric Features	23.00	1.00	+	
Virtual Support Vector Machine	13.00	16.00	+	
Genetic Algorithm	1.80	8.51	+	
Variable Length Segmentation and HMM	4.00	12.00		+
Dynamic Feature of Pressure	6.80	10.80		+
Consistency Functions	1.00	7.00		+
On line SRS - Digitizer Tablet	1.10	3.09		+

8 Conclusions

The Hotelling T^2 statistic presented in this paper is a novel tool for analyzing data. The originality of the proposed approach follows from the fact that classifier utilizes not only extracted features, but also the best similarity measures. In the signature biometrics such an approach has not been applied yet. It can be a good complementary method in biometric investigations.

References

1. Doroz, R., Porwik, P.: Handwritten signature recognition with adaptive selection of behavioral features. In: Chaki, N., Cortesi, A. (eds.) CISIM 2011. CCIS, vol. 245, pp. 128–136. Springer, Heidelberg (2011)
2. Kudłacik, P., Porwik, P.: A new approach to signature recognition using the fuzzy method. Pattern Analysis and Applications (2012), doi:10.1007/s10044-012-0283-9

3. Porwik, P., Para, T.: Some handwritten signature parameters in biometric recognition process. In: Proc. of the 29th IEEE Int. Conf. on Information Technology Interfaces. Dubrovnik, Croatia, pp. 185–190 (2007)
4. Ibrahim, M.T., Kyan, M.J., Guan, L.: On-line signature verification using most discriminating features and Fisher linear discriminant analysis. In: 10th IEEE Int. Symposium on Multimedia, Berkeley CA, pp. 172–177 (2008)
5. Józwik, A., Serpico, S.B., Roli, F.: A parallel network of modified 1-NN and k-NN classifiers-application to remote-sensing image classification. Pattern Recognition Letters 19(1), 57–62 (1998)
6. Lei, H., Govindaraju, V.A.: Comparative study on the consistency of features in on-line signature verification. Pattern Recognition Letters 26(15), 2483–2489 (2005)
7. Kirkwood, B.R., Sterne, J.A.C.: Essentials of medical statistics, 2nd edn. Wiley-Blackwell (2003)
8. Sung-Hyug, C.: Comprehensive survey on distance/similarity measures between probability density functions. Int. J. of Mathematical Models and Methods in Applied Sciences 1, 300–307 (2007)
9. Salvador, S., Chan, P.: Fast DTW: toward accurate dynamic time warping in linear time and space. In: Proc. of the Int. Conf. on Knowledge Discovery and Data Mining, KDD 2004, Seattle, USA, pp. 70–80 (2004)
10. Bharadi, V.A., Kekre, H.B.: Off-line signature recognition systems. Int. J. of Computer Applications 1(27), 48–56 (2010)
11. Impedovo, D., Pirlo, G.: Automatic signature verification: The state of the art". IEEE Trans. on Syst. Man. and Cybernetics. Part C: Applications and Reviews 38(5), 609–635 (2008)
12. Froelich, W., Kisiel-Dorohinicki, M., Nawarecki, E.: Agent-based evolutionary model for knowledge acquisition in dynamical environments. In: International Conference on Computational Science, vol. (3), pp. 839–846 (2006)
13. Koprowski, R., Wróbel, Z.: The cell structures segmentation. In: 4th International Conference on Computer Recognition Systems (CORES 2005), pp. 569–576 (2005)
14. <http://www.cse.ust.hk/svc2004>

Part III

Data Stream Classification and Big Data

Analytics

Predictive Regional Trees to Supplement Geo-Physical Random Fields

Annalisa Appice, Sonja Pravilovic, and Donato Malerba

Abstract. Nowadays ubiquitous sensor stations are deployed to measure geo-physical fields for several ecological and environmental processes. Although these fields are measured at the specific location of stations, geo-statistical problems demand for inference processes to supplement, smooth and standardize recorded data. We study how predictive regional trees can supplement data sampled periodically in an ubiquitous sensing scenario. Data records that are similar one to each other are clustered according to a rectangular decomposition of the region of analysis; a predictive model is associated to the region covered by each cluster. The cluster model depicts the spatial variation of data over a map, the predictive model supplements any unknown record that is recognized belong to a cluster region. We illustrate an incremental algorithm to yield time-evolving predictive regional trees that account for the fact that the statistical properties of the recorded data may change over time. This algorithm is evaluated with spatio-temporal data collections.

1 Introduction

Several industrial and scientific applications collect geo-physical data at specific locations (data sources) but demand for predictions (estimations) where no measured values are available. Two prominent examples are the prediction of air pollution and soil moisture in an area. The interpolation theory (see [9] for an exhaustive survey) investigates the mathematical methods to perform this predictive process by standardizing, supplementing and smoothing the sampled data. Several accurate

Annalisa Appice · Sonja Pravilovic · Donato Malerba

Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro via Orabona,

4 - 70126 Bari - Italy

e-mail: {annalisa.appice, donato.malerba}@uniba.it

Sonja Pravilovic

Montenegro Business School, Mediterranean University Vaka Djurovica b.b. Podgorica - Montenegro

e-mail: sonja.pravilovic@uniba.it

interpolators have been reported in the literature. However they are typically based on a lazy learning approach which delay the prediction step until a specific query is made.

In this paper, we pursue a different approach which aims at building a predictive model before a query is made. Given a set of geo-referenced records that are sampled periodically for a number of numeric geo-physical fields (target properties), we first look for a spatial clustering of these records. Then we associate each cluster with a distinct predictive model which is used to predict the target properties for any unknown record belonging to the cluster. The main benefit of this predictive clustering approach over conventional spatial clustering methods is that they provide both clusters and their symbolic descriptions, as in conceptual clustering [10]. However, unlike conceptual clustering, predictive clustering exhibits a form of supervised learning.

We focus on Predictive Clustering Trees (PCTs) [1], a class of tree structured models, where nodes are associated with clusters and leaves with a single predictive model for multiple target properties. In particular, we consider Predictive Regional Trees (PRT), which are PCTs specifically developed for spatial domains. The top-node of a PRT corresponds to the entire region under examination and contains all the sampled records. This region is recursively partitioned along the spatial coordinates (e.g. latitude and longitude) into smaller sub-regions. In a PRT a predictive model (the mean) is computed for each target property and then associated to each cluster sub-region.

A PRT provides us a static model of some geo-physical phenomenon. Nevertheless, the temporal dimension of the data streams generated by sensors demands for dynamic models, which can be updated continuously as new fresh data arrive. In this work, we propose an incremental algorithm for the construction of time-adaptive PRTs. When a new sample of records (snapshot) is acquired thought stations of a sensor network, a past predictive clustering model is modified to model data of the process that may change their properties (e.g. mean and variance) over time. In theory, a distinct tree can be learned for each time point and several trees can be subsequently combined using some general framework (e.g. [15]) for tracking cluster evolution over time. However, this solution is prohibitively time-consuming when data arrive at high rate. By taking into account that (1) a geo-physical field is often slowly time-varying, and (2) a change of the properties of the data distribution of a field is often restricted to a delimited area of the network, more efficient learning algorithms can be derived. In this work we propose an algorithm which retains the predictive clustering as long as it discriminates between regions of spatial correlated data, while it mines a novel predictive clustering only if the latest becomes inaccurate. In this way, we save computation time without loss of predictive accuracy and track the evolution of the cluster model by detecting changes in data properties at no additional cost.

The paper is organized as follows. In Section 2, we revise related works. In Section 3 we describe the algorithm to mine PRTs while in Section 4 we explain the incremental version for time-adapting PRTs. Finally, in Section 5 we present the evaluation of the incremental strategy and draw conclusions.

2 Related Work

The basics of this study comes from both the interpolation theory and the research literature on incremental learning of predictive models.

The interpolation theory has a long history so that several interpolation algorithms, such as inverse distance weighting [13], Kriging [12], polynomial regression [2] and trend surfaces [17]), have been reported in the literature. Inverse distance weighting and Kriging use a linear combination of weights at sampled points to supplement records at any unknown point. They are both accurate interpolators, especially Kriging, but they are based on a lazy learning approach where estimation is delayed until a query is made for a specific location. Polynomial regression and trend surface analysis implement a general inference phase to reveal the functional model of spatial variation of data. However, different fields originate separate inference processes. Moreover, no incremental update of the functional model has been investigated yet.

The seminal work that addresses the problem of incremental learning of regression trees is presented in [11]. It describes an incremental method to grow a linear model tree (trees with linear models at leaves) by processing one-by-one the examples of a training set. But this work neither addresses a task of multi-target regression (several properties are predicted in a single task) nor considers the temporal information of records coming with an evolving data stream. More recently, an incremental algorithm to learn regression trees from time-changing data streams is presented in [7]. This incremental strategy, also fine-tuned to multi-target problems [6], is defined for general purpose streams. Records are consecutive; once again they are processed one by one, but in the order of arrival. This scenario is still different from that considered in this study where a sample of geo-referenced records is, in general, collected at the same time point. This scenario suggests that the incremental algorithm should be triggered on a new sample of equally-timestamped records (rather than on a single record).

3 Predictive Regional Clustering

Let \mathbf{Z} be the space of m numeric geo-physical fields Z_1, Z_2, \dots, Z_m (target properties); \mathbf{R} be the region of the analysis (also called sampling region); and K be the set of n training stations installed at specific positions of \mathbf{R} . A station of K is geo-referenced in \mathbf{R} through geographic coordinates (e.g. latitude and longitude). A training sample $D(\mathbf{Z}, K)$ is a set of records of measurements for \mathbf{Z} , which are acquired by the stations of K at a specific time point. The *regional clustering task* aims at partitioning \mathbf{R} in sub-regions by maximizing similarity between records of D clustered in the same region. The *predictive task* aims at estimating the (unknown) target properties of \mathbf{Z} for any random position over \mathbf{R} .

The PRTs learner addresses both the clustering task and predictive task. It assigns the role of descriptive variable for the clustering task to the geographic coordinates

Algorithm 1 function $\text{prt}(D)$

Require: D : the snapshot of data collected at the present time point

```

1: if  $\text{stop}(D)$  then
2:   return  $(\text{Var}(D), \text{leaf}(D))$ 
3: else
4:    $(c^*, \text{Var}^*, \mathcal{P}^*) \leftarrow (\text{null}, \text{inf}, \emptyset)$ 
5:   for each possible Boolean test  $c$  on  $K$  do
6:      $\mathcal{P} \leftarrow \{D_L, D_R\}$                                  $\triangleright$  clustering induced by  $c$  on  $D$ 
7:      $\text{Var}(\mathcal{P}) \leftarrow \left( \frac{|D_L| \text{Var}(D_L) + |D_R| \text{Var}(D_R)}{|D|} \right)$ 
8:     if  $(\text{Var}(\mathcal{P}) < \text{Var}^*)$  then
9:        $(c^*, \text{Var}^*, \mathcal{P}^*) \leftarrow (c, \text{Var}(\mathcal{P}), \mathcal{P})$ 
10:    end if
11:   end for
12:   return  $(\text{Var}^*, \text{tree}(c^*, \text{prt}(D_L^*), \text{prt}(D_R^*)))$ 
13: end if

```

received with K . It uses this descriptive variable to partition \mathbf{R} in (a hierarchy of) regions that contain data records similar one to each other. Each region of clustering is associated to a predictive model for each target property of \mathbf{Z} . The conjunction of tests across the hierarchy to the cluster describes the conditions according to we are able to establish if a (virtual) station is installed into a cluster region and use the predictive model of the cluster to estimate the record of its unknown measurements of \mathbf{Z} .

The algorithm (see Algorithm 1) is based on that presented in [1]. Its peculiarity is the definition of a multi-coordinate splitting test that involves the geographic coordinates of a record. The algorithm takes as input the training set D and splits it recursively until a stopping criterion is satisfied. The main loop (Alg. 1, ln 5-11) looks for the best Boolean test c^* for the internal node t and calls itself recursively to construct a sub-tree for each sub-cluster of the partition set $\mathcal{P}^*(t)$ induced by c^* on $D(t)$ (Alg. 1, ln 12). $D(t)$ denotes the set of training records of D falling in t . A candidate test c is in the form:

$$\bigwedge_{C_i \in \text{GeographicCoordinates}} C_i \leq \alpha_i, \quad (1)$$

where α is a threshold between adjacent bins of the equal-width discretization of values that the geographic coordinate C_i assumes for stations of K clustered in t . It is noteworthy that this discretization process to determine the candidate thresholds is performed on-line on each internal node of the tree. In this way, independently on the number of discretization bins, finer-grain thresholds are considered by traversing the tree from top to bottom. The best split test is that which minimizes the average intra-cluster data variance (Var in Alg. 1, ln 7). This variance heuristic is computed on the target properties of records in $D(t)$. Let $\mathcal{P}(t) = \{D_L(t), D_R(t)\}$ be the partition set of $D(t)$ (Alg. 1, ln 6) to be evaluated, $\text{Var}(\mathcal{P}(t))$ is computed as follows,

$$Var(\mathcal{D}(t)) = \left(\frac{|D_L(t)|Var(D_L(t)) + |D_R(t)|Var(D_R(t))}{|D_L(t)| + |D_R(t)|} \right), \quad (2)$$

where $|\cdot|$ denotes the cardinality of a set, while $Var(\cdot)$ is the average value of variance taken over the set of target properties, that is,

$$Var(D(t)) = \frac{\sum_{Z_i \in \mathbf{Z}} variance(scaledZ_i, D(t))}{|D(t)|}. \quad (3)$$

To compute Equation 3, values of each target property are scaled between 0 and 1 such that each target property contributes equally to the overall $Var(\cdot, \cdot)$ measure. The scaling operation to compute $Var(D(t))$ is performed on $D(t)$; scaled values of target properties are used for the computation of the variance heuristic only.

The search stops in two cases (Alg. 1, In 1). The former occurs when the node hosts a number of records that is smaller than $2\sqrt{n}$. This is a good locality threshold presented in [5]. The latter, already used in traditional regression tree learners, occurs when that the data variance computed on training records falling in the node is below a user-defined percentage (ε) of the data variance computed on the entire training set. If a stopping condition is satisfied, a leaf is created (Alg. 1, In 2); the prototype (arithmetic mean for every target property) of records clustered in the leaf is computed. It is noteworthy that, like for traditional regression trees, the use of the mean as a predictive model associated to a cluster is consistent with the variance heuristic according to clusters are determined.

4 Incremental Algorithm

A spatial-primitive strategy is used to process a stream of geo-physical data. Measured records are segmented into consecutive time stamped snapshots such that each snapshot collects the sample of records measured at a specific time point. The snapshots are mined one by one in the order of their arrival. Each time a new snapshot is acquired, the predictive clustering model is modified to model appropriately the present snapshot.

Algorithm 2 describes an incremental algorithm that looks for regions revealing a substantial changes in the properties (mean, variance) of the clustered data with respect to that observed in the past. It learns again parts (i.e. sub-trees) of the existing tree that neither cluster nor predict appropriately the new data. The algorithm inputs a new snapshot D and the previous PRT tree. It prunes a node of $tree$ if the clustering associated with its split does not perform a real reduction of the data variance on the new snapshot; it grows new (sub-)trees of clusters; it modifies a past predictive model in order to fit the statistical properties of the new data in D . The algorithm traverses the complete tree in a post-order; it applies the pruning operator bottom-up. Let t be the visited node, $D(t)$ be the set of records of D which fall in t , the algorithm distinguishes between two cases.

Algorithm 2 $\text{tict}(tree, D)$

Require: $tree$: the tree computed at the previous time point
Require: D : the snapshot of data measure collected at the present time point

```

1:  $t \leftarrow \text{root}(tree)$ 
2: if  $\text{isLeaf}(t)$  then
3:   if  $(\text{stop}(D))$  then
4:     updatePredictiveModel( $t, D$ )
5:     return ( $\text{Var}(D)$ , new  $tree(t)$ )
6:   else
7:     return  $\text{prt}(D)$ 
8:   end if
9: else
10:   $c \leftarrow \text{test}(t)$                                  $\triangleright$  split condition at the node t
11:   $\mathcal{P}_D \leftarrow \{D_L, D_R\}$                        $\triangleright$  clustering induced by  $c$  on  $D$ 
12:   $(\text{Var}_L, tree_L) \leftarrow \text{iprt}(\text{leftTree}(tree), D_L)$ 
13:   $(\text{Var}_R, tree_R) \leftarrow \text{iprt}(\text{rightTree}(tree), D_R)$ 
14:   $\text{Var}(\mathcal{P}) \leftarrow \frac{|D_L(t)|\text{Var}_L + |D_R(t)|\text{Var}_R}{|D_L(t)| + |D_R(t)|}$ 
15:  if  $\text{Var}(D) \times sF > \text{Var}(\mathcal{P})$  then
16:    return ( $\text{Var}(\mathcal{P})$ ,  $tree(c, tree_L, tree_R)$ )
17:  else
18:    return  $\text{prt}(D)$ 
19:  end if
20: end if

```

In the case t is a leaf, the algorithm checks if the leaf condition (one of the stopping criteria described in Section 3) is satisfied on $D(t)$. When this condition is satisfied, the algorithm updates the predictive model at the leaf in order to account for these new data (Alg. 2, lines 3-5). Formally, for each target property Z , t is labeled with the new predictive model $\widehat{Z}(t, time_i)$ that is defined as follows:

$$\widehat{Z}(t, time_i) = \frac{\widehat{Z}(t, time_{i-1}) \times time_{i-1} + \overline{Z}(t, time_i) \times time_i}{time_{i-1} + time_i}, \quad (4)$$

where $\widehat{Z}(t, time_{i-1})$ is the predictive model of Z that labeled t at the time $time_{i-1}$, $\overline{Z}(t, time_i)$ is the arithmetic mean of measurements for Z that falls now in $D(t)$. This decay-time model to down-weight the older data is consistent with a strategy widely adopted across a broad class of data streaming systems, sensor networks, and other distributed monitoring systems that process data with timestamped information [4]. Intuitively, it reflects the belief that the most recent data is most relevant for any analysis, while older data is of less significance, so they are counted, since informative of the process, but at a lower weight. When the leaf condition is not satisfied on $D(t)$, the algorithm looks for specializing the predictive clustering model of $D(t)$.

A new sub-tree rooted in t is learned (Alg. 2, lines 3,7) by trying to minimize the variance heuristic through the clustering of $D(t)$.

In the case t is an internal node, the algorithm computes the variance heuristic on $D(t)$ when the sub-tree rooted in t is kept. This is compared to the measure of data variance made on $D(t)$ when the sub-tree is pruned and a leaf is associated to t . If the maintenance of a sub-tree rooted in t reveals a reduction of the data variance, the sub-tree is maintained in the tree (Alg. 2, lines 15,16). Otherwise, it is discarded and a new sub-tree is learned in its place (Alg. 2, lines 15,18) by clustering $D(t)$. Formally, let $\mathcal{P}(t)$ be the partition set defined on $D(t)$ by the sub-tree rooted in t , this cluster set is inappropriate for $F(t)$ iff $Var(D) \times sF > Var(\mathcal{P})$ with

$$Var(\mathcal{P}) = \frac{\sum_{D_i \in \mathcal{P}} Var(D_i)}{|\mathcal{P}|}$$

$Var(\cdot)$ computed according to Equation 3. The pruning condition multiplies the measure of data variance (that computed in case of pruning) by a scaling factor $sF = (|D(t)| - 1)/(|D(t)| + 1)$. This is inspired by the scaling factor already used to prune regression trees in [16]. In this paper, the scaling factor decreases the effect of a random reduction of the data variance that is in general expected with any data partitioning. In the recursive formulation of the algorithm (see recursive calls of routine iprt() in Alg. 2, lines 12-13) the bottom-up application of pruning operation is applied until further pruning decreased the data variance.

5 Empirical Evaluation

The goal of this empirical study is two-fold. The former is to analyze accuracy and efficiency of the incremental PRTs learner in case of both multi-field data and massive data. The latter is to track the evolution of predictive clusters in an application of real interest for our territory. We use three data collections. The Intel Berkeley Lab (IBL) data set (3000 snapshots) [8] collects measures of multi-target properties (humidity, temperature, voltage and luminosity) every 31 seconds on a network of 54 sensors installed in the Intel Berkeley Lab. South America Air Climate (SAC) data set (372 snapshots) [14] collects monthly-measures of air temperature at 6477 weather stations in South America. PhotoVoltaic Plant (PVP) data set (52 snapshots) [3] collects the weekly electricity production of 52 PhotoVoltaic plants in the South of Italy. For each data collection, the incremental algorithm (IPRT) is compared to the traditional algorithm (PRT) run from scratch at each new snapshot of the stream. The root mean squared error (rmse) for field, the number of end-clusters and the computation time spent for a snapshot are analyzed.

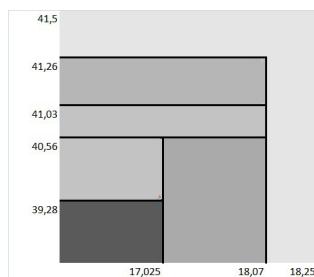
Without loss of generality for the analysis in this paper, PRT and IPRT are run with $\varepsilon = 10\%$. The in-node discretization is performed by looking for twenty bins at each node. For each stream, one half of devices is switched-off so that training is done on the switched-on sensors and testing on the switched-off sensors. Ten random trials of partitioning in training/testing data are generated; results are averaged on these trials.

Table 1 PRT vs IPRT: average root mean squared error (rmse), computation time (millisecs) and number of end clusters. The best rmse is in bold.

Data	Field	IPRT			PRT		
		rmse	time	Nr. of clusters	rmse	time	Nr. of clusters
IBL	Temperature	1.38	4.49	3.33	1.44	7.6	3.28
	Humidity	2.86	-	-	2.79	-	-
	Luminosity	360.62	-	-	366.15	-	-
	Voltage	0.071	-	-	0.072	-	-
SAC	Temperature	3.35	14.7	22.5	2.52	603.129	21.32
PVP	Electricity production	0.92	0.9	3.92	0.95	2.17	4.0

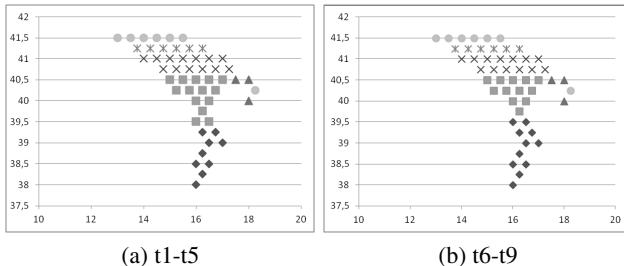
Results reported in Table 1 suggest several considerations. First the incremental algorithm always speeds-up the mining process. In particular, the incremental learning algorithm is able to yield a time-evolving PRT in (near) real time also when data are massive and spread on large networks (see results on SAC). Second, the use of the incremental strategy does not affect greatly the number of regional clusters retrieved for snapshot, while it often brings down the error performed to supplement testing data (see IPRT vs PRT). The exception is SAC where the predictive error is greater with the incremental algorithm. In any case, a deeper analysis reveals that SAC collects measures of temperature that range between -7.6° C and 32.9° C. Based on this information, we judge acceptable that predictive error increases slightly from 2.52 (PRT) to 3.35 (IPRT) if the associated computation time for snapshot decreases greatly from 603.129 msec (PRT) to 14.7 msec (IPRT). Finally, all these considerations on accuracy, number of clusters and computation times are supported equally by the single-field (SAC and PVP) and multi-field (IBL) problems of this study. We complete this study with the analysis of the map of predictive clusters mined for the electricity productions collected in the South of Italy (PVP). For this analysis we consider the predictive clustering model mined from data measured by the entire network of 52 plants. The model of the first snapshot is reported in Figure 1. It reveals a clustering of data performed along the geographic coordinates (see splits on Latitude and Longitude in Figure 1a) that identifies non-overlapping regions in the South of Italy. The picture of these regions on the map confirms the existence of spatial correlation on measures of electricity production over (large) cluster regions, the boundaries between clusters define the change of the properties of data correlation across the space (see regions in Figure 1b). By considering also the predictive model associated to each cluster, we are able to rank these cluster regions based on the electricity production hypothesis. For example, the South of Calabria (diamonds in Figures 1-2.(a)) as well as the South of Apulia (triangles in Figures 1-2.a) are confirmed to be two zones having high production of electricity. Finally, the time-evolution of this map allows us to know the time of any change in the model of electricity production and the arrangement of this change across the space. For example, the time-evolving map reported in Figure 2 shows a cluster model that changes every four-five (weekly) snapshots.

```
+[0-25] Lo<= 18,074995 AND La<= 40,566628
++[0-8] Lo<= 17,025 AND La<= 39,283314
[Electricity Production=16,83889]
++[9-25] Lo> 17,025 ORLa> 39,283314
+++[9-22] Lo<= 17,025 AND La<= 40,566628
[Electricity Production=15,705]
+++[23-25] Lo> 17,025 ORLa> 40,566628
[Electricity Production=16,449999]
+[26-51] Lo> 18,074995 ORLa> 40,566628
++[26-44] Lo<= 18,074995 AND La<= 41,266617
+++[26-38] Lo<= 18,074995 AND La<= 41,033287
[Electricity Production=15,373077]
+++[39-44] Lo> 18,074995 ORLa> 41,033287
[Electricity Production=15,108333]
+++[45-51] Lo> 18,074995 ORLa> 41,266617
[Electricity Production=14,21]
```



(a) predictive regional tree

(b) clustering boundary

Fig. 1 The predictive regional tree at the time point t1 of the PVP stream**Fig. 2** Time-evolving map of predictive clustering on PVP

6 Conclusions

This paper illustrated an incremental algorithm to yield time-evolving predictive regional trees (PRTs) for numeric geo-physical fields. Experiments confirm the efficacy of the describe learner in several spatio-temporal problems. As a future work, we plan to investigate the combination of a predictive regional tree with polynomial regression as well as linear interpolation techniques like inverse distance weighting or Kriging. We also intend to extend the presented algorithm in order to deal with both numeric and discrete fields.

Acknowledgements. This work is carried out in fulfillment of the research objectives of EMP3 Project “Efficiency Monitoring of Photovoltaic Power Plants” funded by the “Fondazione Cassa di Risparmio di Puglia”, and PRIN 2009 Project “Learning Techniques in Relational Domains and Their Applications” funded by the Italian Ministry for Universities and Research (MIUR).

References

1. Blockeel, H., De Raedt, L., Ramon, J.: Top-down induction of clustering trees. In: ICML 1998, pp. 55–63. Morgan Kaufmann (1998)
2. Burrough, P., McDonnell, R.: Principles of Geographical Information Systems. Oxford University Press (1998)
3. Ciampi, A., Appice, A., Malerba, D., Muolo, A.: An intelligent system for real time fault detection in PV plants. In: M'Sirdi, N., Namaane, A., Howlett, R.J., Jain, L.C. (eds.) Sustainability in Energy and Buildings. SIST, vol. 12, pp. 235–244. Springer, Heidelberg (2012)
4. Gama, J., Rodrigues, P.P.: An overview on mining data streams. In: Abraham, A., Hasbanien, A.-E., de Carvalho, A.P.d.L.F., Snášel, V. (eds.) Foundations of Computational Intelligence (6). SCI, vol. 206, pp. 29–45. Springer, Heidelberg (2009)
5. Góra, G., Wojna, A.: RIONA: A classifier combining rule induction and k-NN method with automated selection of optimal neighbourhood. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) ECML 2002. LNCS (LNAI), vol. 2430, pp. 111–123. Springer, Heidelberg (2002)
6. Ikonomovska, E., Gama, J., Dzeroski, S.: Incremental multi-target model trees for data streams. In: SAC 2011, pp. 988–993. ACM (2011)
7. Ikonomovska, E., Gama, J., Dzeroski, S.: Learning model trees from evolving data streams. Data Min. Knowl. Discov. 23(1), 128–168 (2011)
8. Intel Berkeley Labdata, <http://db.csail.mit.edu/labdata/labdata.html>
9. Li, J., Heap, A.: A Review of Spatial Interpolation Methods for Environmental Scientists. Geoscience Australia, Record 2008/23 (2008)
10. Michalski, R.S., Stepp, R.E.: Learning from observation: Conceptual clustering. In: Machine Learning: An Artificial Intelligence Approach, Tioga, pp. 331–364 (1983)
11. Potts, D., Sammut, C.: Incremental learning of linear model trees. Machine Learning 61(1-3), 5–48 (2005)
12. Shekhar, S., Chawla, S.: The origins of kriging. Mathematical Geology 22, 239–252 (1990)
13. Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: ACM National Conference, pp. 517–524. ACM (1968)
14. South American Air Climate Data, http://climate.geog.udel.edu/climate/html_pages/sa_air_clim.html
15. Spiliopoulou, M., Ntoutsi, I., Theodoridis, Y., Schult, R.: Monic: modeling and monitoring cluster transitions. In: Proc. of KDD 2006, pp. 706–711. ACM (2006)
16. Wang, Y., Witten, I.: Induction of model trees for predicting continuous classes. In: Proc. Poster Papers of ECML 1997, pp. 128–137. Faculty of Informatics and Statistics, University of Economics, Prague (1997)
17. Watson, G.: Trend-surface analysis. Journal of the International Association for Mathematical Geology 3, 215–226 (1971)

Extending Bagging for Imbalanced Data

Jerzy Błaszczyński, Jerzy Stefanowski, and Łukasz Idkowiak

Abstract. Various modifications of bagging for class imbalanced data are discussed. An experimental comparison of known bagging modifications shows that integrating with undersampling is more powerful than oversampling. We introduce Local-and-Over-All Balanced bagging where probability of sampling an example is tuned according to the class distribution inside its neighbourhood. Experiments indicate that this proposal is competitive to best undersampling bagging extensions.

1 Introduction

Class imbalance is one of obstacles for learning accurate classifiers. Standard learning algorithms tend to show a strong bias toward the majority classes and misclassify too many examples from the minority class. Several methods to address class imbalance have been proposed so far (see, e.g., [8] for a review). In general, they are categorized in *data level* and *algorithm level* ones. Methods within the first category try to re-balance the class distribution inside the training data by either adding examples to the minority class (*oversampling*) or removing examples from the majority class (*undersampling*). The other category covers methods modifying the learning algorithm, its classification strategy or adapting the problem to cost sensitive framework. New type of ensembles of component classifiers is also visible among these methods. They modify sampling strategies (e.g., in bagging), integrate the ensemble with specialized pre-processing method (e.g. SMOTEBagging [13]) or use different cost weights in generalizations of boosting (see, e.g., [7]).

Although these ensembles are presented as a remedy to a certain imbalanced problems, there is still a lack of a wider study of their properties. Authors often

Jerzy Błaszczyński · Jerzy Stefanowski · Łukasz Idkowiak
Institute of Computing Science, Poznań University of Technology, ul. Piotrowo 2,
60–965 Poznań, Poland
e-mail: {jurek.blaszczyński, Jerzy.Stefanowski}@cs.put.poznan.pl

compare their proposals against the basic versions of other methods. The set of considered imbalance data is usually also limited. Up to now, only two studies were carried out in different experimental frameworks [7, 11]. In [7] authors presented a wide study of 20 different ensembles (all with C4.5 tree classifiers) over 44 data sets. They considered a quite wide range of ensembles from simple modifications of bagging to complex changes of cost Adaboost or hybrid approaches. Their main conclusions said that SMOTEBagging, RUBoost and RUBagging presented the best AUC results. It was also shown that simple versions of undersampling or SMOTE re-sampling combined with bagging worked better than more complex solutions. In [11], two best boosting and bagging ensembles were compared over noisy and imbalanced data. Different amounts of class noise were randomly injected in the training data. The experimental results showed that bagging significantly outperformed boosting. The difference was more significant when data were more noisy. Another surprising conclusion said that it was better to implement sampling without replacement in bagging.

We focus our interest on bagging extensions for class imbalance - following both these related works and its potential usefulness for better handling massive data streams than more complex ensembles. We want to study behavior of bagging extensions more precisely than in [7, 11]. In particular, roughly balanced bagging [9] was missed in [7], although it is appreciated in the literature. The study presented in [11], was too much oriented to the noise level and only two versions of random undersampling in bagging were considered.

Our first objective is to study more precisely a wider set of known extensions of bagging. We will consider exactly balanced bagging and rough balanced bagging but also more variants of using oversampling in bagging – including new type of integrating SMOTE method. We want to check experimentally whether undersampling is better than oversampling in extended bagging. The other contribution is to introduce new extensions of bagging. They are based on the analysis of local neighborhood of each example, which affect the probability of its selection into bootstrap sample – which is a different perspective than the known integrations with pre-processing methods.

2 Adapting Bagging for Imbalanced Data

Bagging introduced by Breiman [4] is an ensemble of T base classifiers induced by the same learning algorithm from T bootstrap samples drawn from the original training set. The predictions of component classifiers form the final decision as the result of equal weight majority voting. The key concept is *bootstrap* aggregation, where the training set for each classifier is constructed by random uniformly sampling (with replacement) instances from the original training set (usually keeping the size of the original data).

2.1 Related Modifications of Bagging

The majority of proposals modify the bootstrap sampling by integrating it with data pre-processing. As the original data is imbalanced the bootstrap sampling will not change significantly the class distribution in the training sample. Its bias toward the majority class could be changed in different ways.

Exactly Balanced Bagging (**EBBag**) is based on a kind of undersampling, which reduces the number of the majority class examples in each bootstrap sample to the cardinality of the minority class (S_{min}) [5]. While creating each training sample, the entire minority class is copied and combined with randomly chosen subsets of the majority class.

Roughly Balanced Bagging (**RBBag**) results from a critique of EBBag [9]. Instead of fixing the sample size, it equalizes sampling probability of each class. Size of the majority class in each bootstrap sample (BS_{maj}) is determined probabilistically according to the negative binomial distribution. Then, S_{min} , and BS_{maj} examples are drawn with or without replacement from the minority class and majority class, respectively. The class distribution in the bootstrap samples maybe slightly imbalanced and varies over iterations. Authors of [9], said that RBBag is more consistent with the nature of bagging and performs better than EBBag. However, authors of [11] claim that there is no significant difference in performance between RBBag and EBBag.

Another way to overcome class imbalance in the bootstrap sampling is oversample the minority class. OverBagging (**OverBag**) is the simplest version which applies oversampling to create each training bootstrap sample. S_{maj} of minority class examples is sampled with replacement to exactly balance the cardinality of the minority and the majority class in each sample. Majority examples are sampled with replacement as in the original bagging.

Another methodology is used in SMOTEBagging (**SMOTEBag**) to increase diversity of component classifiers [13]. Firstly, SMOTE is used instead of random oversampling of the minority class. Then, the SMOTE resampling rate (α) is step-wise changed in each iteration from small to high values (e.g., from 10% till 100%). This ratio defines the number if minority examples ($\alpha \times S_{min}$) to be additionally re-sampled in each iteration. Quite similar trick to construct bootstrap training samples is also used in the "from undersampling to oversampling" ensemble. According to [7] SMOTEBag gave slightly better results than other good random undersampling ensembles.

2.2 New Proposals

We propose to consider another approaches to bagging based on analyzing the local characteristics of examples and focusing sampling toward more consistent ones. Following [12], type of an example (with respect to its classification consistency) can be identified by analyzing its local neighborhood. It can be modeled by class

assignments of its k -nearest neighbours. For each example, the ratio n/k representing the consistency weight is defined, where n is the number of neighbor examples that have the same class label (the Heterogeneous Valued Distance Metric is used and k is relatively small $k=5$).

Firstly, we consider modifications of Roughly Balanced Bagging, where the probabilities of sampling the majority class examples are changed. Instead of using equal probabilities they are tuned to reflect the consistency weight of examples. We consider two variants. In **RBBagV1**, weights of the examples correctly re-classified by their neighbours are set to 1 (to focus more interest on safer regions while reducing the role of border examples). On the other hand, in **RBBagV2** the probability of drawing the example is proportional to n/k instead of equal probabilities. In both variants, weights of outlier examples (with all neighbours from the other class) are set to 0.1. The minority examples are drawn with uniform probabilities (as in RBBag).

In the other approach, we prioritize changing probability of example drawing, without modifying the size of the bootstrap samples. In this way, we drop the idea to integrate either undersampling nor oversampling and stay with the original bootstrap idea. Our intuition is still to focus the sampling toward more safe, consistent examples. The approach also results in two variants. In Over-All Balanced Bagging (**O-ABBag**), we perform overall (global) balancing. For each majority example, its weight is reduced according to proportion of majority class in the original data, i.e., $\frac{S_{min}+S_{maj}}{S_{maj}}$. The weights of minority examples are increased analogously,

with respect to proportion $\frac{S_{min}+S_{maj}}{S_{min}}$. Local-and-Over-All Balanced Bagging (**L-O-ABBag**), on the other hand, uses a mixture of local and global balancing. For each example, its weight is tuned, w.r.t. (with respect to) O-ABBag, in the way to reflect the *local imbalance*, i.e., imbalance in its neighborhood. The tuning takes a form of product of the local and global proportions under the assumption that the local imbalance is independent on the global imbalance. The first component of the product is global proportion, which is calculated exactly as in O-ABBag. The second component is the local proportion that has the same mathematical form as global proportion, however it is calculated taking into account only the examples from neighborhood. The neighborhood is composed of all similar examples, i.e., examples, which share the same description on nominal features and have the value in the same interval on numerical ones (i.e., local discretization) [3]. We consider neighborhoods which are constructed on random subsets of the original feature set (see [2]). More precisely, for each example, we calculate the local proportion w.r.t. random subset of features of size equal to ln of the feature set size. This idea is unlike the previously considered k -nearest neighbours. Instead of taking given k similar examples w.r.t. all features, we take all similar examples, however, w.r.t small random subset of features. In this way, the local proportion promotes examples that are distinct (having different description than any other example) or that have the common description for the class. Taking random subsets of features is a way to consider larger neighborhoods, which may help to obtain more diversified bootstrap samples used in bagging.

3 Experiments

In the first experiments we compare literature best extensions of bagging, while in the second experiments we evaluate our new extensions proposed in the previous section. All implementations are done for WEKA framework. Component classifiers b are learned with C4.5 tree learning algorithm (J4.8), which uses standard parameters except disabling pruning.

We choose 22 real-world data sets representing different domains, sizes and imbalance ratio. Most of them come from the UCI repository and are often used in other works on class imbalance, e.g. in [1]. Other data sets come from our medical applications. For data sets with more than two decision classes, we chose the smallest class as a minority class and aggregated other classes into one majority class.

The performance of classifiers is measured using: *sensitivity* of the minority class, its *specificity* and their aggregation to the *geometric mean* (G-mean). For their definitions see, e.g. [8, 10]. They are estimated in stratified 10-fold cross-validation repeated several times to reduce variance.

3.1 Comparison of Known Bagging Extension

The following bagging variants are considered: Exactly Balanced Bagging (denoted further as EBBag), Roughly Balanced Bagging (RBBag) as the best representatives of undersampling generalization. OverBagging (OverBag) and SMOTEBagging (SMOBag) for oversampling perspectives. In case of using SMOBag, we used 5 neighbours and oversampling ration α was stepwise changed in each sample starting from 10%. Moreover, we decided to use SMOTE in yet another way. In the new ensemble, called BagSMOTE, the bootstrap samples were drawn in a standard way, and than SMOTE was applied to balance majority and minority class distribution in each sample (but with the same α , invariant between samples). For all bagging variants, we tested the following numbers T of component classifiers: 20, 50 and 100. Due to space limit, we present detailed results for $T = 50$ only. Results for other T lead to similar general conclusions. The average values of the sensitivity measure are presented in Table 1. The last row contains averaged ranks calculated as in the Friedman test [10]. The test with post-hoc analysis (the critical difference $CD = 1.61$) shows that EBBag and RBBag leads to significantly better sensitivity than all other bagging variants. However, the more precised Wilcoxon test (with $\alpha=0.05$) shows that differences between these two classifiers are not significant.

While using SMOTE to oversample the minority class, the new integration BagSMOTE performs better than the previously known SMOTE+Bag and OverBag. We also analysed sampling with or without replacement. Conclusions are not univocal. For best undersampling variants like EBBag differences are insignificant while for oversampling standard replacement sampling works much better.

Similar analysis is performed for G-mean (we have to skip details). All extensions are significantly better than the standard version and the ranking of best

Table 1 Sensitivity [%] for known bagging extensions

data set	Bag	EBBag	RBBag	OverBag	SMOBag	BagSMOTE
abdominal pain	72.05 (5)	81.65 (1)	80.08 (2)	74.22 (4)	71.57 (6)	76.86 (3)
acl-m	83.33 (6)	94.17 (1)	88.5 (2)	84.17 (5)	85.0 (4)	88.33 (3)
balance-scale	0 (6) (6)	49.33 (1)	44.2 (2)	8.83 (3)	1.0 (4)	0.67 (5)
breast-cancer	35.93 (5)	56.06 (2)	56.25 (1)	44.91 (3)	34.36 (4)	50.05 (6)
breast-w	94.88 (6)	96.01 (2)	96.27 (1)	95.84 (3)	95.02 (4)	95.17 (5)
bupa	60.48 (5)	66.97 (3)	68.49 (1)	63.27 (4)	57.02 (6)	67.21 (2)
car	73.97 (6)	100 (1.5)	100 (1.5)	92.62 (3)	92.54 (4)	92.13 (5)
cleveland	9.72 (6)	77.22 (1)	73.5 (2)	16.11 (5)	20.83 (4)	36.11 (3)
cmc	36.67 (6)	66.61 (1)	63.62 (2)	46.47 (4)	40.05 (5)	53.10 (3)
german credit	48.89 (5)	72.5 (2)	91.67 (1)	52.89 (4)	45.89 (6)	63.11 (3)
ecoli	56.67 (6)	78.2 (1.5)	78 (1.5)	60.85 (5)	71.67 (4)	77.11 (3)
flags	0 (6)	70 (1)	67.4 (2)	65.27 (3)	55.6 (4)	20 (5)
haberman	26.38 (6)	60.56 (2)	58.39 (3)	49.86 (4)	48.91 (5)	66.25 (1)
hepatitis	49.44 (6)	81 (2)	81.5 (1)	61.67 (3)	54.44 (5)	67.25 (4)
ionosphere	81.79 (6)	85.73 (2)	85.86 (1)	84.7 (3)	83.7 (4.5)	83.76 (4.5)
new-thyroid	87.5 (6)	95.5 (1.5)	95.5 (1.5)	93.06 (4)	92.22 (5)	93.89 (3)
pima	61.28 (6)	76.7 (1)	75.64 (2)	67.38 (3)	65.13 (4)	63.38 (5)
scrotal pain	58.11 (6)	73.78 (2)	74.6 (1)	65.89 (4)	73.8 (3)	58.56 (5)
solar-flareF	7 (6)	86 (2)	86.7 (1)	42.3 (3)	37.33 (4)	34.4 (5)
transfusion	34.62 (6)	65.45 (2)	64.36 (3)	61.88 (4)	56.54 (5)	68.66 (1)
vehicle	91.29 (6)	91.16 (2)	96.78 (1)	93.46 (4)	92.14 (5)	94.97 (3)
yeast-M2	32.22 (6)	90.22 (1)	89.8 (2)	39.9 (5)	41.18 (4)	57.94 (3)
average rank	5.54	1.57	1.52	3.54	4.34	3.52

performing classifiers is the same as for the sensitivity. Again, undersampling methods EBBag and RBBag are better than oversampling bagging variants. However, RBBag seems to be slightly better than EBBag and this trend is more visible for a higher number of component classifiers and using bootstrap sampling with replacement. Bag+SMOTE is also the best performing classifier among oversampling variants. For EBBag and RBBag, we calculated F-measure as yet another popular measure. In this case, RBBag with replacement is better than EBBag with in the Wilcoxon test.

For all bagging variants average values of Q statistics are also calculated to analyze the diversity of component classifiers. Generally, its values are high positive which indicates that diversity is not high. Sampling with replacement improves the diversity. RBBag produces more diverse classifiers than EBBag.

3.2 Newly Proposed Extensions

Then, we compare newly proposed extensions of bagging for class imbalance (see Section 2.2) to Roughly Balanced Bagging, which show good properties. We will consider: two modifications of RBBag (RBBagV1 and RBBagV2), Over-All

Balanced Bagging (O-ABBag), and Local-and-Over-All Balanced Bagging (L-O-ABBag). In case of L-O-ABBag, a random subset of features of size ln of the number of features in the data set is used.

Table 2 Sensitivity [%] calculated for newly proposed extensions of bagging

data set	RBBag	RRBBagV1	RRBBagV2	O-ABBag	L-O-ABBag
abdominal pain	80 (4)	82.26 (1)	82.08 (2)	75.35 (5)	80.4 (3)
acl-m	88.5 (3.5)	89.5 (2)	90 (1)	86 (5)	88.5 (3.5)
balance-scale	44.2 (2)	25.1 (4)	32 (3)	8.163 (5)	99.59 (1)
breast-cancer	56.56 (2)	56.06 (3)	54.97 (4)	51.76 (5)	63.29 (1)
breast-w	96.27 (3)	97.11 (2)	97.52 (1)	95.35 (5)	96.27 (4)
bupa	68.49 (2)	68.37 (3)	66.39 (4)	65.66 (5)	71.03 (1)
car	100 (2.5)	100 (2.5)	100 (2.5)	94.78 (5)	100 (2.5)
cleveland	73.5 (1)	66.67 (4)	67.5 (3)	32.57 (5)	70.29 (2)
cmc	63.62 (2)	61.41 (4)	61.88 (3)	48.41 (5)	68.23 (1)
german credit	89.33 (3)	91 (2)	91.67 (1)	62.6 (5)	71.8 (4)
ecoli	78 (2)	64 (4.5)	64 (4.5)	71.43 (3)	93.14 (1)
flags	66 (4)	67.4 (2)	67.27 (3)	51.76 (5)	92.94 (1)
haberman	58.39 (4)	57.58 (5)	58.58 (3)	59.75 (2)	89.14 (1)
hepatitis	81.5 (1)	76.17 (3)	78.67 (2)	64.38 (5)	71.25 (4)
ionosphere	85.86 (1.5)	85.86 (1.5)	85.4 (3)	85.4 (4)	84.76 (5)
new-thyroid	95.5 (2)	93.33 (3)	92.67 (5)	93.14 (4)	96 (1)
pima	75.64 (4)	76.7 (3)	77.38 (2)	75.3 (5)	79.18 (1)
scrotal pain	74.6 (2)	73.8 (3.5)	73.8 (3.5)	71.53 (5)	77.63 (1)
solar-flareF	86.7 (1)	86.3 (2)	85.4 (3)	41.4 (5)	83.72 (4)
transfusion	64.36 (2)	61.88 (4)	63.12 (3)	60.56 (5)	89.66 (1)
vehicle	96.78 (4)	97.38 (2)	97.29 (3)	93.97 (5)	97.99 (1)
yeast-M2	89.8 (2)	82.33 (4)	83.93 (3)	41.18 (5)	92.94 (1)
average rank	2.48	2.95	2.84	4.68	2.05

Again we show results only for $T = 50$ classifiers. The average values of sensitivity and G-mean are presented in Table 2, and Table 3, respectively (with average ranks). The results of Friedman tests (with CD = 1.3), reveals that only O-ABBg is the worst classifiers. Still, we can give some more detailed observations. According to the average ranks on sensitivity the best performing is L-O-ABBag. However, according to Wilcoxon, its difference to RBBag is not so significant (p -value is just at 0.05). The two modified versions of RBBag perform similarly to each other.

For G-mean, RBBag is the best classifier according to average ranks. However, as Wilcoxon test indicates, its results are not significantly better than these of RB-BagV1, and L-O-ABBag. The worst classifier in comparison is again O-ABBag.

Table 3 G-mean [%] calculated for newly proposed extensions of bagging

data set	RBBag	RRBBagV1	RRBBagV2	O-ABBag	L-O-ABBag
abdominal pain	81.04 (1)	80.78 (2)	80.56 (4)	79.3 (5)	80.73 (3)
acl-m	88.97 (3)	89 (2)	89.24 (1)	87.68 (5)	88.65 (4)
balance-scale	51.32 (1)	35.65 (4)	43 (2)	25.42 (5)	39.8 (3)
breast-cancer	60.28 (1)	60.14 (2)	59.68 (4)	59.53 (5)	59.7 (3)
breast-w	96.12 (4)	96.53 (2)	96.71 (1)	95.73 (5)	96.47 (3)
bupa	71.97 (2)	72.25 (1)	70.67 (3)	70.31 (4)	69.45 (5)
car	96.81 (4)	96.94 (2)	96.78 (5)	96.93 (3)	97.32 (1)
cleveland	73.33 (1)	71.23 (3)	71.14 (4)	53.56 (5)	72.06 (2)
cmc	65.29 (2)	65.25 (3)	65.68 (1)	60.12 (5)	64.18 (4)
german credit	87.07 (3)	88.61 (2)	88.71 (1)	67.67 (5)	67.94 (4)
ecoli	71.84 (3)	63.95 (4.5)	63.95 (4.5)	81.19 (2)	89.04 (1)
flags	67.23 (4)	68.81 (2)	68.6 (3)	63.8 (5)	74.04 (1)
haberman	64.17 (1)	63.04 (4)	63.29 (3)	63.53 (2)	50.74 (5)
hepatitis	80.29 (1)	77.49 (3)	78.85 (2)	74.2 (5)	75.04 (4)
ionosphere	90.75 (5)	90.87 (2)	90.84 (3)	90.79 (4)	90.91 (1)
new-thyroid	96.21 (2)	95.07 (4)	94.34 (5)	95.65 (3)	96.72 (1)
pima	74.84 (3)	75.67 (2)	75.83 (1)	74.44 (4)	73.91 (5)
scrotal pain	74.43 (1)	73.62 (4)	73.41 (5)	74.37 (2)	74.16 (3)
solar-flareF	85.03 (2)	85.05 (1)	84.43 (3)	61.33 (5)	82.12 (4)
transfusion	67.65 (2)	67.82 (1)	67.52 (3)	66.93 (4)	44.77 (5)
vehicle	95.23 (2)	94.9 (3)	94.81 (4)	94.77 (5)	96.2 (1)
yeast-M2	85.57 (1)	82.94 (4)	83.64 (3)	63.32 (5)	85.06 (2)
average rank	2.23	2.61	2.98	4.23	2.95

4 Discussion

The results of first experiments clearly show that applying simple random undersampling leads to much better classification performance than oversampling variants. Definitely, both EBBag and RBBag achieve the best results. However, the difference between them and the best oversampling bagging is much higher than shown in [7]. Moreover, according to our results, SMOTEBagging is not as accurate as it has been presented in [13]. A new oversampling bagging variant, where SMOTE is applied with the same oversampling ratio, works better than the previously promoted SMOTEBagging.

Although EBBag and RBBag performs similarly with respect to the sensitivity, RBBag seems to be slightly better than EBBag for G-mean and F-measure, in particular when sampling is done with replacement. This is a bit different observation to [11], where both classifiers worked similarly on all artificially modified noisy data. Authors of RBBag also showed its slightly better performance over EBBag [9] over 9 data sets only (4 of them was also used in our experiments). Yet another novel observation is that sampling with replacement may be profitable for RBBag unlike

EBBag, where our results show no differences between sampling with or without replacement. This result is contradictory to a quite strong claim, from [11], that “bagging should be implemented without replacement”.

Discussing these results, we can hypothesize that undersampling may help in random distributing majority examples among many small bootstrap samples, which may direct learning to some useful complementary classification patterns. Even though some of bootstrap samples may contain unsafe, difficult majority examples, the final voting with better components reduces their influence. We plan to carry our future experiments of studying the content and diversity of bootstraps in both EBBag and RBBag.

The main methodological contribution of our study is a new extension of bagging called Local-and-Over-All Balanced Bagging. It is based on different principles than all known bagging extensions. Firstly, instead of integrating bagging with pre-processing, we keep the standard bagging idea but we change radically probabilities of sampling examples by increasing the chance of drawing more minority examples. In this sense we somehow oversample some minority examples and go toward another distribution balance inside the bootstrap sample. Moreover, we promote sampling more safe examples with respect to analyzing class distribution in their neighborhood. The next novel contribution is modeling this nearest distribution by finding similar examples with respect to subsets of features describing them. This idea is inspired by our earlier proposal of variable consistency sampling [2] and according to our best knowledge has not been considered yet in case of imbalanced data.

The results of the second part of experiments clearly show that this novel proposal leads to competitive results to best known undersampling bagging extensions. Furthermore, using more local information about class imbalance is more powerful than using only global imbalance ratio (which was often considered in earlier works). In our future works we plan to study more precisely other ways of modeling and using local approaches to class imbalance.

We would also like to extend our comparison to other component classifiers. In final remarks we refer to problems of using these bagging variants to massive data. In our opinion, more complex solutions do not work better than these simpler bagging variants. Notice that bagging is relatively easy to implement (even in a parallel way) and to generalize, e.g., for class imbalance. Therefore, randomly balanced variants of bagging could be attractive with respect to computational costs since appropriate redistribution of the majority class leads to many smaller training samples. Moreover, according to [7], all these bagging extensions lead to smaller trees than boosting variants.

References

1. Batista, G., Prati, R., Monard, M.: A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter 6(1), 20–29 (2004)

2. Błaszczyński, J., Słowiński, R., Stefanowski, J.: Feature Set-based Consistency Sampling in Bagging Ensembles. In: Proc. From Local Patterns To Global Models (LEGO), ECML/PKDD Workshop, pp. 19–35 (2009)
3. Błaszczyński, J., Słowiński, R., Stefanowski, J.: Variable Consistency Bagging Ensembles. *Transactions on Rough Sets* 11, 40–52 (2010)
4. Breiman, L.: Bagging predictors. *Machine Learning* 24(2), 123–140 (1996)
5. Chang, E.: Statistical learning for effective visual information retrieval. In: Proc. of ICIP 2003, pp. 609–612 (2003)
6. Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16, 341–378 (2002)
7. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F.: A Review on Ensembles for Class Imbalance Problem: Bagging, Boosting and Hybrid Based Approaches. *IEEE Transactions on Systems, Man, and Cybernetics—Part C* 42(4), 463–484 (2011)
8. He, H., Garcia, E.: Learning from imbalanced data. *IEEE Transactions on Data and Knowledge Engineering* 21(9), 1263–1284 (2009)
9. Hido, S., Kashima, H.: Roughly balanced bagging for imbalance data. *Statistical Analysis and Data Mining* 2(5-6), 412–426 (2009)
10. Japkowicz, N., Shah, M.: Evaluating Learning Algorithms. A Classification Perspective. Cambridge University Press (2011)
11. Khoshgoftaar, T., Van Hulse, J., Napolitano, A.: Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Transactions on Systems, Man, and Cybernetics—Part A* 41(3), 552–568 (2011)
12. Napierska, K., Stefanowski, J.: Identification of different types of minority class examples in imbalanced data. In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B. (eds.) HAIS 2012, Part II. LNCS, vol. 7209, pp. 139–150. Springer, Heidelberg (2012)
13. Wang, S., Yao, T.: Diversity analysis on imbalanced data sets by using ensemble models. In: Proc. IEEE Symp. Comput. Intell. Data Mining, pp. 324–331 (2009)

Rule Chains for Visualizing Evolving Fuzzy Rule-Based Systems

Sascha Henzgen, Marc Strickert, and Eyke Hüllermeier

Abstract. Evolving fuzzy systems are data-driven fuzzy (rule-based) systems supporting an incremental model adaptation in dynamically changing environments; typically, such models are learned on a continuous stream of data in an online manner. This paper advocates the use of visualization techniques in order to help a user gain insight into the process of model evolution. More specifically, *rule chains* are introduced as a novel visualization technique for the inspection of evolving Takagi-Sugeno-Kang (TSK) fuzzy systems. To show the usefulness of this techniques, we illustrate its application in the context of learning from data streams with temporal concept drift.

1 Introduction

Visualization has become an important tool in data-driven research fields, such as machine learning and data mining. Apart from the visualization of data objects, data relationships, and aggregated information content [4], the visualization of models learned from data has recently attracted increasing attention in the field of computational intelligence [1, 2, 8, 9]. Going beyond the presentation of static models, this paper suggests the use of visualization techniques for tracking *evolving* models, that is, models that are learned and adapted in an online manner on a continuous stream of data [3, 10].

More specifically, we developed an interactive visualization tool that allows for monitoring evolving TSK fuzzy inference systems in real time. So-called *rule chains* are proposed for visualizing changes of a rule system between two consecutive time points. Experimentally, we show that characteristic patterns emerge in a rule chain visualization when applied to evolving rule systems learned from data streams exhibiting concept drift.

Sascha Henzgen · Marc Strickert · Eyke Hüllermeier

Department of Mathematics and Computer Science, Philipps University Marburg, Germany
e-mail: {henzgen, strickert, eyke}@mathematik.uni-marburg.de

This paper is structured as follows. Before introducing our rule chain visualization technique in Section 3, we revisit the underlying TSK fuzzy rule-based systems in Section 2. In Section 4, an experimental study with synthetic data is presented, showing the usefulness of our visualization technique for analyzing the evolution of fuzzy models learned on data streams with concept drift. Conclusions are drawn in Section 5.

2 TSK Fuzzy Rule-Based Systems

In this work, we consider first order Takagi-Sugeno-Kang (TSK) fuzzy rule-based systems $\mathfrak{R} = \{R_1, \dots, R_N\}$ implementing maps of the form

$$f : \mathbb{R}^p \rightarrow \mathbb{R}, x \mapsto \sum_{i=1}^N l_i(x) \cdot \Psi_i(x) . \quad (1)$$

Here, the antecedent of each rule R_i is modeled as a conjunction of p fuzzy sets with Gaussian membership function, each characterized by its center $c_{i,j}$ and width $\sigma_{i,j}$; for a given input vector $x = (x_1, \dots, x_p) \in \mathbb{R}^p$, the relevance (“firing strength”) of an antecedent part is evaluated as follows:

$$\Psi_i(x) = \frac{\exp\left[-\frac{1}{2}\sum_{j=1}^p \left((x_j - c_{i,j})^2 / \sigma_{i,j}^2\right)\right]}{\sum_{k=1}^{|\mathfrak{R}|} \exp\left[-\frac{1}{2}\sum_{j=1}^p \left((x_j - c_{k,j})^2 / \sigma_{k,j}^2\right)\right]} \quad (2)$$

The conclusion of R_i is a linear function specified by a $(p+1)$ -dimensional weight vector $w_i = (w_{i,0}, w_{i,1}, \dots, w_{i,p})$:

$$l_i(x) = w_{i,0} + w_{i,1}x_1 + w_{i,2}x_2 + \dots + w_{i,p}x_p \quad (3)$$

Learning TSK models of that kind on a continuous (and potentially unbounded) stream of data $(z^{(1)}, z^{(2)}, z^{(3)}, \dots)$ in the form of input/output tuples $z^{(t)} = (x^{(t)}, y^{(t)})$ essentially means applying a learning algorithm \mathcal{A} that adapts the current rule model after each newly observed example. Thus, starting with a model $\mathfrak{R}^{(0)}$ at time $t = 0$, a corresponding sequence of models $(\mathfrak{R}^{(0)}, \mathfrak{R}^{(1)}, \mathfrak{R}^{(2)}, \dots)$ is produced, where $\mathfrak{R}^{(t)} = \mathcal{A}(\mathfrak{R}^{(t-1)}, z^{(t)})$ is obtained by modifying certain rules in $\mathfrak{R}^{(t-1)}$, by merging different rules into a single new rule, or by creating a new rule from scratch. For a proper handling of rules, it is important that all rules have a unique label, no matter whether freshly generated or originating from a merging process. A rule $R^{(t)}$ in $\mathfrak{R}^{(t)}$ which evolved from a rule $R_i^{(t-1)}$ in $\mathfrak{R}^{(t-1)}$ with label i keeps that label. The history \mathfrak{H}_i is the set of all rules labeled with i .

3 Visualizing Model Evolution with Rule Chains

In this section, we introduce our idea of rule chains for visualizing important aspects of the evolution of fuzzy rule-based systems. Essentially, a rule chain seeks to capture the changes of a single rule between two consecutive time points. An important prerequisite for visualizing such changes is the availability of meaningful measures of similarity or distance between fuzzy rules. Therefore, prior to explaining our visualization technique, we introduce measures of this kind. Since an independent treatment of the antecedent parts and the conclusion parts of fuzzy rules is informative for different aspects of the visualization, we measure the similarity of both parts separately.

3.1 Similarity and Distance Measures for Fuzzy Rules

3.1.1 Antecedence Similarity

Recall that the antecedence part M_k of a rule R_k is a conjunction of fuzzy sets $\mu_{k,i}$ with normalized Gaussian membership function, one for each input variable x_i (cf. Section 2). We define the similarity between two antecedence parts M_k and M_l of two rules R_k and R_l by

$$\mathfrak{S}(M_k, M_l) = \min \left(\mathfrak{s}(\mu_{k,1}, \mu_{l,1}), \mathfrak{s}(\mu_{k,2}, \mu_{l,2}), \dots, \mathfrak{s}(\mu_{k,p}, \mu_{l,p}) \right), \quad (4)$$

where $\mathfrak{s}(\mu_{k,i}, \mu_{l,i})$ is a standard similarity between fuzzy sets, namely the size of their intersection (pointwise minimum of membership degrees) normalized by the size of the larger of the two:

$$\mathfrak{s}(\mu_{k,i}, \mu_{l,i}) = \frac{|\mu_{k,i} \cap \mu_{l,i}|}{\max(|\mu_{k,i}|, |\mu_{l,i}|)} \quad (5)$$

The size $|\mu|$ of a Gaussian fuzzy set μ is defined by the area under the membership function. Thus, its computation comes down to solving an integration problem for which no closed-form solution exists. Therefore, we exploit the connection between the cumulative distribution function of the normal distribution and the error function $\text{erf}(z) = 2/\sqrt{\pi} \int_0^z e^{-\tau^2} d\tau$, for which efficient numerical implementations exist. Thus, the area under a normalized Gaussian function with mean c and standard deviation σ can be efficiently determined as follows:

$$FN_{c,\sigma}(z) = \frac{\sigma\sqrt{2\pi}}{2} \left(1 + \text{erf} \left(\frac{z-c}{\sigma\sqrt{2}} \right) \right) \quad (6)$$

3.1.2 Distance between the Rule Centers

Another similarity measure based on rule antecedence parts is the distance between rule centers, where the center of a rule R_k is defined as

$$c_k = (c_{k,1}, c_{k,2}, \dots, c_{k,p}) , \quad (7)$$

with $c_{k,i}$ the center of the i^{th} fuzzy set in the antecedence of the k^{th} rule. By using the Euclidean metric, the distance between two rules is

$$\mathfrak{D}(R_k, R_l) = \|c_k - c_l\| . \quad (8)$$

For the purpose of visualization, it is desirable to have the measure normalized to the range $[0, 1]$. In rule chains, the computation of distances is restricted to adjacent rules $R_i^{(t)}$ and $R_i^{(t+1)}$ from the same history \mathfrak{H}_i (the history of the t^{th} rule). Therefore, the normalization is done as follows:

$$\mathfrak{D}\mathfrak{N}\left(R_i^{(t)}, R_i^{(t+1)}\right) = \frac{\mathfrak{D}\left(R_i^{(t)}, R_i^{(t+1)}\right)}{\max_{(R_i^{(\tau)}, R_i^{(\tau+1)}) \in \mathfrak{H}_i \times \mathfrak{H}_i} \mathfrak{D}\left(R_i^{(\tau)}, R_i^{(\tau+1)}\right)} \quad (9)$$

3.1.3 Angle Similarity

The angle similarity is a measure of similarity between the conclusion parts of two rules. The conclusion part (3) of a TSK fuzzy rule R_i defines a p -dimensional hyperplane $H_i = \{(x_1, x_2, \dots, x_n, l_i(x)) | x \in \mathbb{R}^n\}$ with normal vector $v_i = (w_{i,1}, w_{i,2}, \dots, w_{i,p}, -1)$. Using this normal vector, we can compute the angle between two conclusions H_i and H_j by

$$\alpha = \arccos\left(\frac{v_i \cdot v_j}{\|v_i\| \cdot \|v_j\|}\right) , \quad (10)$$

and finally their angle similarity [6, 7] as

$$\mathfrak{S}_\alpha(H_i, H_j) = \begin{cases} 1 - \frac{2}{\pi}\alpha & \text{if } \alpha \leq \frac{\pi}{2} \\ \frac{2}{\pi}(\alpha - \frac{\pi}{2}) & \text{else} \end{cases} . \quad (11)$$

3.1.4 Conclusion Intercept Difference

In addition to the angle similarity of two conclusions, we also consider the difference in their y -intercept, which is not captured by this similarity. Although this difference might indeed be negligible on a global scale, one should keep in mind that the influence of a conclusion is localized by the rule antecedence; and locally, the constant term may clearly make a difference. Therefore, we define the conclusion y -intercept difference by

$$\mathfrak{C}\mathfrak{V} \left(R_i^{(t)}, R_i^{(t+1)} \right) = \frac{\left| w_{i,0}^{(t)} - w_{i,0}^{(t+1)} \right|}{\max_{(R_i^{(\tau)}, R_i^{(\tau+1)}) \in \mathfrak{H}_i \times \mathfrak{H}_i} \left| w_{i,0}^{(\tau)} - w_{i,0}^{(\tau+1)} \right|}, \quad (12)$$

with $w_{i,0}$ denoting the y -intercept of the conclusion of the rule R_i .

3.2 Rule Chains

We propose rule chains as an adequate means for visualizing specific aspects of the evolution of a fuzzy rule-based system. A rule history \mathfrak{H} is visualized as a horizontal pearl chain, where every pearl represents a rule at a certain time point. The first time point is located at the left and the most recent time point at the right end. Pearls along a chain are connected with different types of links. Focusing on a measure of interest, links between adjacent pearls can represent the corresponding similarity or distance between consecutive rules. As illustrated in Figure 1, there are three types of links that correspond, respectively, to the measures introduced in Section 3.1.

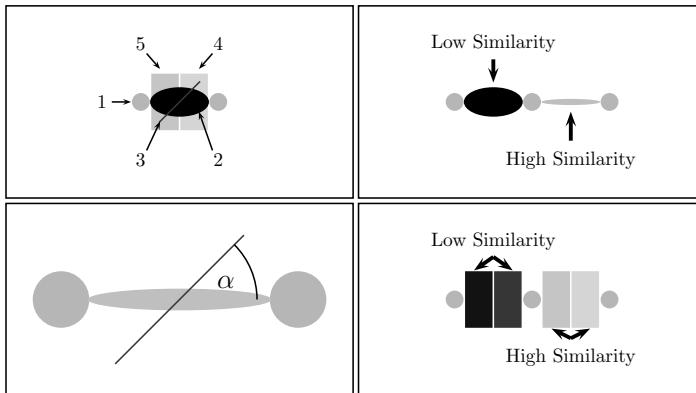


Fig. 1 The upper left panel shows all five elements of a rule chain. The elements pointed at by the numbered arrows are: (1) rule pearl, (2) antecedence similarity link, (3) angle line, (4) angle similarity block, (5) conclusion shift block. The top right panel describes the meaning of different heights and color intensities of ellipsoid links. The bottom right panel describes the meaning of different color intensities of rectangular links. Generally, both types of links can be used for visualizing any given similarity or distance measure. The lower left panel shows the rotated line segment for visualizing the angle between the conclusions of two adjacent rules.

In addition to the horizontal time point position, the vertical position of a pearl determines the history of the associated rule. The antecedence similarity link corresponds to the antecedence similarity measure or the Euclidean distance of centers. The degree of similarity is connected to the color intensity and shape of the ellipse. The lower the similarity, the bigger and darker is the ellipse. The angle similarity

block visualizes the angle similarity, and the conclusion shift block displays differences of y -intercepts of the related rule conclusions. The color intensity of both blocks is the higher, the lower the similarity and the higher the difference is, respectively. Thus, the three types of links highlight different aspects of change between two rules. The angle line is mathematically positively rotated by α against the horizon to display the angle between two temporally adjacent conclusions.

For visualizing the evolving system as a whole, the horizontal rule chains are stacked (Figure 2). A single rule chain gives a quick overview of a rule's lifetime and development, and stacking them provides a holistic view on the development of the whole rule chain system.

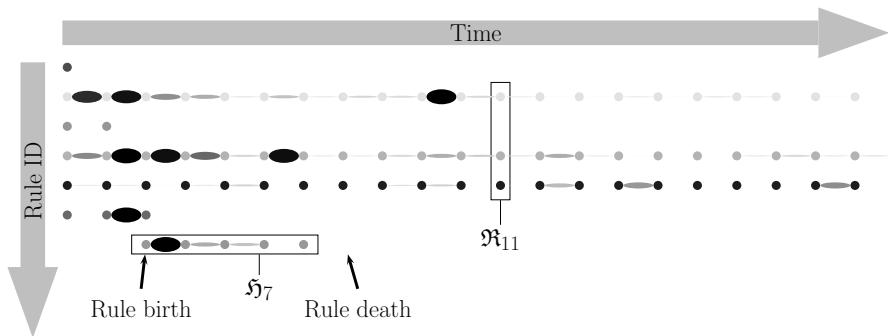


Fig. 2 This figure shows a cut-out from a rule chain system. All rules, symbolized by pearls, belonging to the same rule system are aligned vertically. Rules belonging to the same rule history are organized in horizontal chains. The pearl links are associated with similarity measures. Here, only the antecedence similarity links (ellipses) are shown.

3.3 Concept Drift Detection with Rule Chains

An important requirement for evolving (fuzzy) systems is a quick discovery and adequate reaction to so-called *concept drift* [3]. Roughly speaking, a concept drift is a (gradual) change of the data-generating process in the course of time, that is, a change of the probability distribution P that generates input/output tuples emitted by the data stream.

Rule chains provide an adequate (visual) means for discovering concept drift and monitoring the fuzzy system's reaction to this drift. In fact, as a reaction to a change of the data generating process, the fuzzy system is expected to adapt a possibly large number of rules (both, antecedence and/or conclusion parts). Likewise, new rules will typically be created and existing ones will be deleted.

Changes of that kind naturally produce observable patterns in the visualized rule chain system. More precisely, a simultaneous change of the antecedence or conclusion parts of many rules will produce noticeable vertical lines in the rule chain system (Figure 3). Such patterns can be amplified by defining thresholds for the

minimal change to be drawn. The simultaneous appearance or disappearance of many rules produces long vertical edges in the displayed rule chains. This enables the user to recognize potential concept drifts in the data. More correctly, since we are visualizing the model and not the data evolution, the observation of such patterns should only be taken as an indication (and not as a proof) of a possible concept drift.

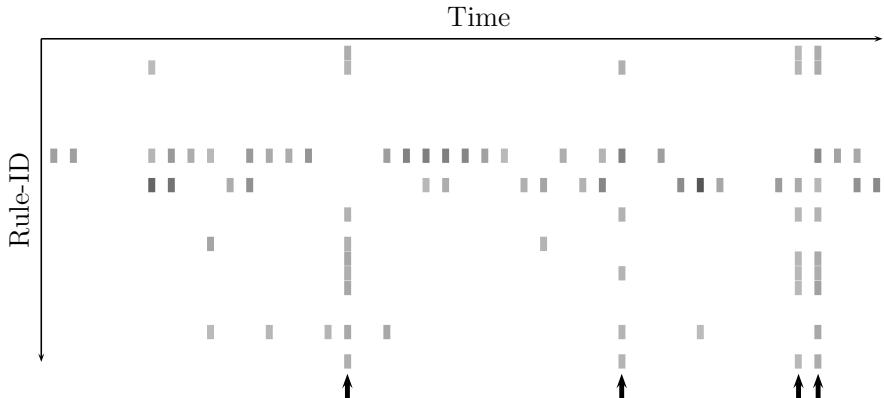


Fig. 3 This figure highlights patterns in the rule chains which indicate significant changes in the model and, therefore, suggest a possible concept drift in the data. The bold arrows are marking time points at which an abrupt change of the conclusion parts of many rules can be observed as vertical lines.

4 Illustration

This section presents an illustration of our rule chain visualization tool that is meant as a first proof of concept. For this purpose, we make use of synthetic data, which allows for conducting controlled experiments and, since the “ground truth” is known for this data, for judging the plausibility of the results—properties that are obviously not offered by real data. Moreover, while our visualization tool is in principle independent of the learning algorithm, we used FLEXFIS [5] as a concrete implementation and set the forgetting factor to 0.9. This FLEXFIS specific factor controls the forgetting of the inverse hessian matrix during the recursive weighted least squares optimization of the rule consequents.

To generate a data stream, we concatenate a number of synthetic datasets. Every dataset is associated with a time point determined by the position p of the dataset in the concatenation. An instance of a dataset consists of a two-dimensional input and a real-valued output. For the initial dataset, three normally distributed clusters C_1, C_2, C_3 are generated. Every cluster C_i is characterized by its center $c_i = (c_{i,1}, c_{i,2})$, its width $\sigma_i = (\sigma_{i,1}, \sigma_{i,2})$ and the size n_i . The output y associated with an instance $x = (x_1, x_2)$ depends on the cluster C_i and is determined by a linear function:

$$y = f_i(x) = w_{i,0} + w_{i,1}(x_1 - c_{i,1}) + w_{i,2}(x_2 - c_{i,2}) . \quad (13)$$

Thus, a dataset can be characterized by $(p, (c_i, \sigma_i, n_i, w_i)_{i=1}^m)$, where p is the position of the dataset in the stream and m is the cluster number. For the whole stream, only a few datasets lying on so-called anchor positions are explicitly declared. The datasets lying between two anchor positions are interpolated by a linear function (i.e., the parameters c, σ, n, w characterizing a data set are convex combinations of the corresponding parameters of the left and right anchor datasets). The characteristics of the data stream used in our study are summarized in Table 1.

Table 1 Specification of the data stream. Cluster size of each cluster at any time is $n = 1000$. The cluster width is $\sigma_1 = 2$ and $\sigma_2 = 2$ for all clusters.

p	C_1		C_2		C_3		C_1, C_3			C_2		
	c_1	c_2	c_1	c_2	c_1	c_2	w_0	w_1	w_2	w_0	w_1	w_2
0	10	50	50	50	90	50	0	10	0	0	10	0
3	10	50	50	50	90	50	0	10	0	0	10	0
6	10	50	50	50	90	50	-10	0	0	10	0	0
12	10	50	50	50	90	50	-10	0	0	10	0	0
16	10	-100	50	150	90	-100	-10	0	0	10	0	0
26	10	-100	50	150	90	-100	-10	0	0	10	0	0
30	10	-100	50	150	90	-100	-10	0	10	10	0	10

The data stream contains three concept drifts with two stable episodes in-between, namely 6–12 and 16–26. These episodes allow the learner to re-stabilize the model. At the beginning, all three clusters are well separated and located next to each other along the first dimension. Moreover, the three regression functions share the same coefficients. Somewhat surprisingly, this initial configuration lets FLEXFIS generate more than only the expected three rules and leads to rich patterns during model formation.

The first drift is caused by a change of the regression values (between anchor positions 3 and 6, i.e., time steps 3000 and 6000). In Figure 4, this drift phase lies between marks A and B. At mark 1, a reaction of FLEXFIS is clearly seen. The second drift (between anchor positions 12 and 16) is caused by a movement of the clusters, which are all moving in different directions. The period of this drift is delimited by marks C and D. The visible reaction of FLEXFIS is between marks 2 and 3. Here, a change in the antecedent parts of the rules is seen, which confirms the expectation from the experimental design. The last concept drift (between anchor positions 26 and 30) is again caused by a drift in the regression values. It starts at mark E and ends at F. A change in the rule system caused by this drift can be seen from mark 4 onwards. After the third drift, the rule system does not reach a stable stage until the end of the experiment. All three concept drifts cause an adaptation of the rule system by the learning algorithm, resulting in clearly visible patterns in the rule chain visualization.

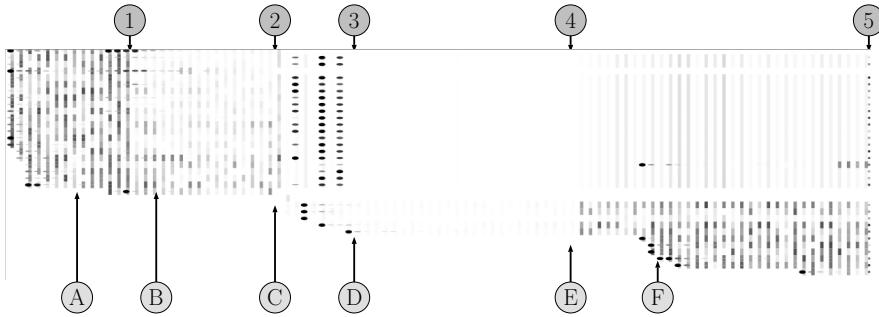


Fig. 4 This figure shows the result of a visualization of an evolving fuzzy rule-based system using stacked rule chains. There are three concept drifts in the underlying data stream, the first between A and B, the second between C and D, and the third between E and F. The numbers mark the time points where a change in the rule system, as a reaction to the concept drift, becomes visible.

5 Conclusion

We proposed stacked pearl chains for visually tracking the development of rules in evolving TSK fuzzy rule-based systems. Not only the life span of individual rules can be assessed by looking at the rule history, also demographic changes of the evolving rule system become visible. Temporal rule chaining is accomplished by visualizing similarities between adjacent rule antecedents and conclusions. Basically, the overlap between evolved rules or their center proximities is being displayed, as well as the angle between their conclusions.

With an example application using synthetic data, we highlighted several benefits of our visualization tool: the size of the rule system (number of rules), being a good indicator of model complexity, can be grasped quickly and monitored over time quite easily. Moreover, vertical patterns clearly indicate concept drifts and corresponding reactions of the evolving model.

Rule chain visualization as introduced in this paper is part of an ongoing software project for interactive monitoring of evolving TSK model evolution. This way, insights into model dynamics can be attained, which is interesting for both end-users who receive early feedback about setting model-specific parameters and developers who like to study convergence or adaptation properties of their models. Future work is focusing on the display of rule similarities and evolving rules at attribute level, global model state characteristics, and interactive control of evolving fuzzy system models. Eventually, our tool is supposed to provide the user access to complex states in time-adaptive machine learning methods beyond evolving TSK models.

Acknowledgements. This project is supported by the German Research Foundation (DFG). We also thank the LOEWE Center for Synthetic Microbiology (SYNMIKRO), Marburg, for financial support.

References

1. Alonso, J.M., Cordón, O., Quirin, A., Magdalena, L.: Analyzing interpretability of fuzzy rule-based systems by means of fuzzy inference-grams. In: World Congress on Soft Computing (2011)
2. Gabriel, T.R., Thiel, K., Berthold, M.R.: Rule visualization based on multi-dimensional scaling. In: 2006 IEEE International Conference on Fuzzy Systems, pp. 66–71. IEEE (2006)
3. Gama, J.: A survey on learning from data streams: current and future trends. *Progress in Artificial Intelligence* 1(1), 45–55 (2012)
4. Keim, D.A., Kohlhammer, J., Ellis, G., Mansmann, F.: Mastering The Information Age - Solving Problems with Visual Analytics. In: Eurographics (2010)
5. Lughofer, E.: FLEXFIS: A robust incremental learning approach for evolving Takagi-Sugeno fuzzy models. *IEEE Transactions on Fuzzy Systems* 16(6), 1393–1410 (2008)
6. Lughofer, E., Hüllermeier, E.: On-line redundancy deletion in evolving fuzzy regression models using a fuzzy inclusion measure. In: Galichet, S., Montero, J., Mauris, G. (eds.) Proc. Eusflat–2011, 7th Int. Conf. of the European Soc. for Fuzzy Logic and Technology, pp. 380–387 (2011)
7. Lughofer, E., Bouchot, J.-L., Shaker, A.: On-line elimination of local redundancies in evolving fuzzy systems. *Evolving Systems* 2(3), 165–187 (2011)
8. Peters, G., Bunte, K., Strickert, M., Biehl, M., Villmann, T.: Visualization of processes in self-learning systems. In: Tenth Annual International Conference on Privacy, Security and Trust (TSOS), pp. 244–249 (2012)
9. Rehm, F., Klawonn, F., Kruse, R.: Rule classification visualization of high-dimensional data. In: Proc. of the 11th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems, IPMU 2006 (2006)
10. Sayed-Mouchaweh, M., Lughofer, E.: Learning in non-stationary environments. Springer (2012)

Recovery Analysis for Adaptive Learning from Non-stationary Data Streams

Ammar Shaker and Eyke Hüllermeier

Abstract. The extension of machine learning methods from static to dynamic environments has received increasing attention in recent years; in particular, a large number of algorithms for learning from so-called *data streams* has been developed. An important property of dynamic environments is *non-stationarity*, i.e., the assumption of an underlying data generating process that may change over time. Correspondingly, the ability to properly react to so-called *concept change* is considered as an important feature of learning algorithms. In this paper, we propose a new type of experimental analysis, called *recovery analysis*, which is aimed at assessing the ability of a learner to discover a concept change quickly, and to take appropriate measures to maintain the quality and generalization performance of the model.

1 Introduction

The development of methods for learning from so-called *data streams* has been a topic of active research in recent years [6, 9]. Roughly speaking, the key idea is to have a system that learns incrementally, and maybe even in real-time, on a continuous and potentially unbounded stream of data, and which is able to properly adapt itself to changes of environmental conditions or properties of the data generating process. Systems with these properties have already been developed for different machine learning and data mining tasks, such as clustering and classification [7].

An extension of data mining and machine learning methods to the setting of data streams comes with a number of challenges. In particular, the standard “batch mode” of learning, in which the entire data as a whole is to provided as an input to the learning algorithm (or “learner” for short), is no longer applicable. Correspondingly, the learner is not allowed to make several passes through the data set, which is commonly done by standard methods in statistics and machine learning. Instead,

Ammar Shaker · Eyke Hüllermeier

Department of Mathematics and Computer Science, University of Marburg, Germany
e-mail: {shaker, eyke}@mathematik.uni-marburg.de

the data must be processed in a single pass, which implies an incremental mode of learning and model adaptation.

Domingos and Hulten [3] list a number of properties that an ideal stream mining system should exhibit, and suggest corresponding design decisions: the system uses only a limited amount of memory; the time to process a single record is short and ideally constant; the data is volatile and a single data record accessed only once; the model produced in an incremental way is equivalent to the model that would have been obtained through common batch learning (on all data records so far); the learning algorithm should react to concept change (i.e., any change of the underlying data generating process) in a proper way and maintain a model that always reflects the current concept.

This last property is often emphasized as a key feature of learning algorithms, since non-stationarity is arguably the most important difference between static and dynamic environments. Indeed, while the idea of an incremental learning is crucial in the setting of data streams, too, it is not entirely new and has been studied for learning from static data before. The ability of a learner to maintain the quality and generalization performance of the model in the presence of concept drift, on the other hand, is a property that becomes truly important when learning under changing environmental conditions.

In this paper, we propose a new type of experimental analysis, called *recovery analysis*, which is aimed at assessing this ability of a learner. Roughly speaking, recovery analysis suggests a specific experimental protocol and a graphical presentation of a learner's generalization performance that provides an idea of how quickly a drift is recognized, to what extent it affects the prediction performance, and how quickly the learner manages to adapt its model to the new condition.

2 Learning under Concept Drift

We consider a setting in which an algorithm \mathcal{A} is learning on a time-ordered stream of data $S = (z_1, z_2, z_3, \dots)$. Since we are mainly interested in *supervised learning*, we suppose that each data item z_t is a tuple $(x_t, y_t) \in \mathbb{X} \times \mathbb{Y}$ consisting of an input x_t (typically represented as a vector) and an associated output y_t , which is the target for prediction. In classification, for example, the output space \mathbb{Y} consists of a finite (and typically small) number of class labels, whereas in regression the output is a real number.

At every time point t , the algorithm \mathcal{A} is supposed to offer a predictive model $\mathcal{M}_t : \mathbb{X} \rightarrow \mathbb{Y}$ that has been learned on the data seen so far, i.e., on the sequence $S_t = (z_1, z_2, \dots, z_t)$. Given a query input $x \in \mathbb{X}$, this model can be used to produce a prediction $\hat{y} = \mathcal{M}_t(x) \in \mathbb{Y}$ of the associated output. The accuracy of this prediction can be measured in terms of a loss function $\ell : \mathbb{Y} \times \mathbb{Y} \rightarrow \mathbb{R}$, such as the 0/1 loss in the case of classification or the squared error loss in regression. Then, the prediction performance of \mathcal{M}_t is defined in terms of the expected loss, where the expectation

is taken with respect to an underlying probability measure \mathbf{P} on $\mathbb{Z} = \mathbb{X} \times \mathbb{Y}$. This probability measure formally specifies the data generating process.

If the algorithm \mathcal{A} is truly incremental, it will produce \mathcal{M}_t solely on the basis of \mathcal{M}_{t-1} and z_t , that is, $\mathcal{M}_t = \mathcal{A}(\mathcal{M}_{t-1}, z_t)$. In other words, it does not store any of the previous observations z_1, \dots, z_{t-1} . Most algorithms, however, store at least a few of the most recent data points, which can then also be used for model adaptation. In any case, the number of observations that can be stored is typically assumed to be finite, which excludes the possibility of memorizing the entire stream. A *batch learner* \mathcal{A}_B , on the other hand, would produce the model \mathcal{M}_t on the basis of the complete set of data $\{z_1, \dots, z_t\}$. Note that, although \mathcal{A} and \mathcal{A}_B have seen the same data, \mathcal{A}_B can exploit this data in a more flexible way. Therefore, the models produced by \mathcal{A} and \mathcal{A}_B will not necessarily be the same.

As mentioned before, the data generating process is characterized by the probability measure \mathbf{P} on $\mathbb{Z} = \mathbb{X} \times \mathbb{Y}$. Under the assumptions of stationarity and independence, each new observation z_t is generated at random according to \mathbf{P} , i.e., the probability to observe a specific $z \in \mathbb{Z}$ is given by¹

$$\mathbf{P}(z) = \mathbf{P}(x, y) = \mathbf{P}(x) \cdot \mathbf{P}(y|x) .$$

Giving up the assumption of stationarity (while keeping the one of independence), the probability measure \mathbf{P} generating the next observation may possibly change over time. Formally, we are thus dealing, not with a single measure \mathbf{P} , but with a sequence of measures $(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \dots)$, assuming that z_t is generated by \mathbf{P}_t . One speaks of a *concept change* if these measures are not all equal [1].

In the literature, a distinction is made between different causes and types of concept change [8]. The first type refers to a sudden, abrupt change of the underlying concept to be learned and is often called *concept shift* (\mathbf{P}_t is very different from \mathbf{P}_{t-1}). Roughly speaking, in the case of a concept shift, any knowledge about the old concept may become obsolete and the new concept has to be learned from scratch. The second type refers to a gradual evolution of the concept over time. In this scenario, old data might still be relevant, at least to some extent. Finally, one often speaks about *virtual* concept drift if the change only concerns $\mathbf{P}(x)$, i.e., the distribution of the inputs, while the concept itself, i.e., the conditional distribution $\mathbf{P}(y|x)$, remains unchanged [14]. To guarantee optimal predictive performance, an adaptation of the model might also be necessary in such cases. In practice, virtual and real concept drift will often occur simultaneously.

Learning algorithms can handle concept change in a direct or indirect way. In the indirect approach, the learner does not explicitly attempt to detect a concept drift. Instead, the use of outdated or irrelevant data is avoided from the outset. This is typically accomplished by considering only the most recent data while ignoring older observations, e.g., by sliding a window of fixed size over a data stream. To handle concept change in a more direct way, appropriate techniques for discovering the drift or shift are first of all required, for example based on statistical tests.

¹ We slightly abuse notation by using the same symbol for the joint probability and its marginals.

3 Recovery Analysis

In practical studies, data streams are of course never truly infinite. Instead, a “stream” is simply a large data set in the form of a long yet finite sequence $S = (z_1, z_2, \dots, z_T)$. In experimental studies, such streams are commonly used to produce a performance curve showing the generalization performance of a model sequence $(\mathcal{M}_t)_{t=1}^T$ over time. Although many of these studies are interested in analyzing the ability of a learner to deal with concept drift, such an analysis is hampered by at least two problems: First, for a data stream S , it is normally not known whether it contains any concept drift, let alone when such a drift occurs; this is a problem at least for real data, while obviously less of an issue if data is generated synthetically. Second, even if a concept drift is known to occur, it is often difficult to assess the performance of a learner or to judge how well it recovers after the drift, simply because a proper *baseline* is missing: The performance that could in principle be reached, or at least be expected, is not known.

3.1 Main Idea and Experimental Protocol

In order to overcome these problems, our idea is to work, not with a single data stream, but with three streams in parallel, two “pure streams” and one “mixture”. The pure streams $S_A = (z_1^a, z_2^a, \dots, z_T^a)$ and $S_B = (z_1^b, z_2^b, \dots, z_T^b)$ are supposed to be stationary and generated, respectively, according to distributions \mathbf{P}_A and \mathbf{P}_B ; in the case of real data, stationarity of a stream can be guaranteed, for example, by permuting the original stream at random. These two streams must also be compatible in the sense of sharing a common data space $\mathbb{Z} = \mathbb{X} \times \mathbb{Y}$. The mixture stream $S_C = (z_1^c, z_2^c, \dots, z_T^c)$ is produced by randomly sampling from the two pure streams:

$$z_t^c = \begin{cases} z_t^a & \text{with probability } \lambda(t) \\ z_t^b & \text{with probability } 1 - \lambda(t) \end{cases} \quad (1)$$

A concept drift can then be modeled, for example, by specifying the (time-dependent) sample probability $\lambda(t)$ as a sigmoidal function:

$$\lambda(t) = \left(1 + \exp\left(\frac{t-t_0}{w}\right) \right)^{-1}.$$

This function has two parameters: t_0 is the mid point of the change process, while w controls the length of this process. Using this transition function, the stream S_C is obviously drifting “from S_A to S_B ”: In the beginning, it is essentially identical to S_A , in a certain time window around t_0 , it moves away from S_A toward S_B , and in the end, it is essentially identical to S_B . Thus, we have created a gradual concept drift with a rate of change controlled by w .

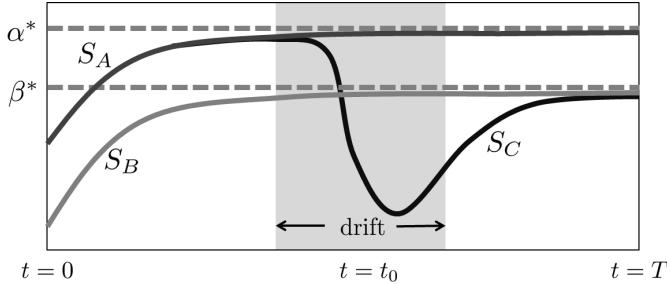


Fig. 1 Schematic illustration of a recovery analysis: The three performance curves are produced by training models on the pure streams S_A and S_B , as well as on the mixed stream S_C , each time using the same learner \mathcal{A} . The region shaded in grey indicates the time window in which the concept drift (mainly) takes place. While the concept is drifting, the performance on S_C will typically drop to some extent.

Now, suppose the same learning algorithm \mathcal{A} is applied to all three streams S_A , S_B and S_C . Since the first two streams are stationary, we expect to see a standard learning curve when plotting the generalization performance (for example, the classification accuracy) as a function of time. In the following, we denote the performance curves for S_A and S_B by $\alpha(t)$ and $\beta(t)$, respectively. These curves are normally concave, showing a significant increase in the beginning before reaching a certain saturation level later on; see Figure 1 for an illustration. The corresponding saturation levels α^* and β^* provide important information, namely information about the best performance that can be expected by the learner \mathcal{A} on the pure streams S_A and S_B , respectively.

Even more interesting, however, is the performance curve $\gamma(t)$ for the stream S_C , which exhibits concept drift. In the beginning, this curve will be effectively identical to the curve for S_A , so that the learner \mathcal{A} should reach the level α^* . Then, upon the beginning of the concept drift, the performance is expected to drop, and this decrease is supposed to continue until the drift ends and the learner \mathcal{A} starts to recover. Eventually, \mathcal{A} may (or may not) reach the level β^* . This level is indeed an upper bound on the asymptotic performance, since \mathcal{A} cannot do better even when being trained on S_B from the very beginning. Thus, reaching this level indicates an optimal recovery.

Obviously, the performance curve for S_C provides important information about the ability of \mathcal{A} to deal with concept drift. In particular, the minimum of this curve indicates how strongly \mathcal{A} is affected by the concept drift. Moreover, the curve informs about how quickly the performance deteriorates (giving an idea of how sensitive \mathcal{A} is), how much time \mathcal{A} needs to recover, and whether or not it manages to recover optimally.

3.2 Bounding the Optimal Generalization Performance

As explained above, the performance curve produced by a learner \mathcal{A} on the stream S_C is expected to decrease while this stream is drifting from S_A to S_B . In order to judge the drop in performance, not only relatively in comparison to other learners but also absolutely, it would be desirable to have a kind of reference performance as a baseline. This leads to an interesting question: Is it possible to quantify our expectations regarding the drop in performance? More specifically, what is the optimal generalization performance

$$\gamma^*(t) = \sup_{\mathcal{M} \in \mathbf{M}} \gamma_{\mathcal{M}}(t) \quad (2)$$

we can expect on the stream S_C at time t ? Here \mathbf{M} is the underlying model class (i.e., the class of models that \mathcal{A} can choose from), and $\gamma_{\mathcal{M}}(t)$ denotes the generalization performance of a model $\mathcal{M} \in \mathbf{M}$ on the mixture distribution (1), i.e.,

$$\mathbf{P}_C(t) = \lambda(t)\mathbf{P}_A + (1 - \lambda(t))\mathbf{P}_B .$$

Our experimental setup indeed allows for answering this question in a non-trivial way. To this end, we exploit knowledge about the performance levels $\alpha(t)$ and $\beta(t)$ that can be reached on S_A and S_B , respectively. Thus, there are models $\mathcal{M}_A, \mathcal{M}_B \in \mathbf{M}$ whose performance is $\alpha_{\mathcal{M}_A}(t) = \alpha(t)$ and $\beta_{\mathcal{M}_B}(t) = \beta(t)$. Now, suppose we were to apply the model \mathcal{M}_A on the stream S_C . What is the expected generalization performance? If an example (x, y) on S_C is generated according to \mathbf{P}_A , the generalization performance (expected loss) of \mathcal{M}_A on this example is the same as on S_A , namely $\alpha_{\mathcal{M}_A}(t)$. Otherwise, if the example is generated according to \mathbf{P}_B , nothing can be said about the performance of \mathcal{M}_A ; thus, assuming that the performance measure takes values in the unit interval, we can only assume the worst case performance of 0. Since the first case occurs with a probability of $\lambda(t)$ and the second one with a probability of $1 - \lambda(t)$, the overall expected performance of \mathcal{M}_A is given by

$$\lambda(t) \cdot \alpha_{\mathcal{M}_A}(t) + (1 - \lambda(t)) \cdot 0 = \lambda(t) \cdot \alpha_{\mathcal{M}_A}(t) .$$

Using the same line of reasoning, the performance of the model \mathcal{M}_B on the stream S_C is given by $(1 - \lambda(t))\beta_{\mathcal{M}_B}(t)$. Thus, choosing optimally from the two candidate models $\{\mathcal{M}_A, \mathcal{M}_B\} \subset \mathbf{M}$, we can at least guarantee the performance

$$\gamma^*(t) = \max \{ \lambda(t) \cdot \alpha_{\mathcal{M}_A}(t), (1 - \lambda(t)) \cdot \beta_{\mathcal{M}_B}(t) \} . \quad (3)$$

Obviously, since the supremum in (2) is not only taken over $\{\mathcal{M}_A, \mathcal{M}_B\}$ but over the entire model class \mathbf{M} , $\gamma^*(t)$ is only a lower bound on the optimal performance $\gamma^*(t)$, that is, $\gamma^*(t) \leq \gamma^*(t)$. We also remark that, if the performance levels $\alpha(t)$ and $\beta(t)$ are already close enough to the optimal levels α^* and β^* , respectively, then (3) can be written more simply as

$$\gamma^*(t) = \max \{ \lambda(t) \cdot \alpha^*, (1 - \lambda(t)) \cdot \beta^* \} . \quad (4)$$

Strictly speaking, this estimation is not correct, since α^* and β^* are only limit values that will not necessarily be attained. Practically, however, this is of no importance, especially since we have to work with estimations of these values anyway.

Finally, we note that of course not all performance measures (loss functions) can naturally be normalized to $[0, 1]$. Especially problematic in this regard are measures that are principally unbounded, such as the squared loss in regression. In such cases, an estimation similar to the one above can nevertheless be derived, provided the worst case performance can be bounded by a constant; this constant will then replace the constant in our estimation, which is simply 0.

3.3 Practical Issues

Our discussion of recovery analysis so far has left open some important practical issues that need to be addressed when implementing the above experimental protocol. An obvious question, for example, is how to determine the generalization performance of a model \mathcal{M}_t (induced by the learner \mathcal{A}) at time t , which is needed to plot the performance curve. First of all, it is clear that this generalization performance can only be estimated on the basis of the data given, just like in the case of batch learning from static data. In the literature, two procedures are commonly used for performance evaluation on data streams:² (i) In the *holdout* approach, the training and the test phase of a learner are interleaved as follows: The model is trained incrementally on a block of M data points and then evaluated (but no longer adapted) on the next N instances, then again trained on the next M and tested on the subsequent N instances, and so forth. (ii) In the *test-then-train* approach, every instance is used for both training and testing. First, the current model is evaluated on the observed instance, and then this instance is used for model adaptation. The evaluation measure in this scenario is updated incrementally after each prediction (sequential evaluation). This approach can also be applied in a chunk mode, where a block of size M (instead of a single instance) is used for evaluation first and training afterwards.

The test-then-train procedure has some advantages over the holdout approach. For example, it obviously makes better use of the data, since each example is used for both training and testing. More importantly, it avoids “gaps” in the learning process: In the holdout approach, \mathcal{A} only learns on the training blocks but stops adaptation on the evaluation blocks in-between. Such gaps are especially undesirable in the presence of a concept drift, since they may bias the assessment of the learner’s reaction to the drift. This is the main reason for why we prefer the test-then-train procedure for our implementation of recovery analysis.

Another practical issue concerns the length of the data streams. In fact, to implement recovery analysis in a proper way, the streams should be long enough, mainly to make sure that the learner \mathcal{A} will saturate on all streams: First, it should reach the saturation levels α^* and β^* on S_A and S_B , respectively. Moreover, the streams

² Both procedures are implemented in the MOA framework [2].

should not end while \mathcal{A} is still recovering on S_C ; otherwise, one cannot decide whether or not an optimal recovery (reaching β^*) is accomplished.

Finally, to obtain smooth performance curves, we recommend to repeat the same experiment with many random permutations S_A and S_B of the original streams, and to average the curves thus produced. Obviously, averaging is legitimate in this case, since the results are produced for the same data generating processes (specified by the distributions \mathbf{P}_A , \mathbf{P}_B and their mixture \mathbf{P}_C).

4 Illustration

This section is meant to illustrate our idea of recovery analysis by means of some practical examples. To this end, we conducted a series of experiments with different classification methods. All algorithms were implemented under MOA [2], except FLEXFIS, which is implemented in Matlab. MOA is a framework for learning on data streams. It includes data stream generators and several classifiers. Moreover, it offers different methods for performance evaluation.

Due to a lack of space, we present results only for a single data set, namely the weather data provided by the National Oceanic and Atmospheric Administration (NOAA).³ Since this data originally contained missing values, we used it in the form as suggested by [5]. It contains eight daily weather measurements, such as temperature, visibility, etc. The goal is to predict whether it will be a rainy day or not. We used this data as a first pure stream S_A and an “inverted copy” as a pure stream S_B . In this copy, we simply inverted the target attribute. Thus, the problem on the mixture stream S_C gradually changes from predicting whether it will be rainy to predicting whether it will not be rainy.

Performance curves were produced using the test-then-train procedure (cf. Section 3.3) in chunk mode (with chunk size 500) and averaging over ten random shuffles of the data. At each point of time, the evaluation curve shows the prediction performance on the most recent chunk.

The results for four different methods are shown in Figure 2: the eFPT method for learning evolving fuzzy pattern trees [12], the instance-based learner IBLStreams [11], the Hoeffding Trees classifier for learning decision trees [4] in its incremental form without any specific reaction mechanism to concept changes, and the FLEXFIS method for fuzzy rule induction [10].

From this figure, which also indicates the range of the drift and shows the lower bound on the optimal performance (3) as a reference, some interesting observations can be made. For example, by comparing the lower bound with the respective “performance valleys” of the methods, it can be seen that all methods are reacting with a certain delay, which is clearly expected; some methods, however, are obviously a bit faster than others. The same remark applies to the extent of performance loss, which is, for example, more pronounced for FLEXFIS than for eFPT. Overall, the instance-based learner IBLStreams seems to perform best, showing a curve that is close to the

³ http://users.rowan.edu/~polikar/research/NIE_data/

reference (3). The Hoeffding Tree learner, on the other hand, is performing much worse. Its loss in performance is higher, and it recovers only very slowly. In fact, the decision trees produced by this learner are often very complex and, therefore, difficult to adapt if significant changes are needed. This may explain why Hoeffding Trees react more slowly than other learners (a tendency that we could also confirm on other data streams).

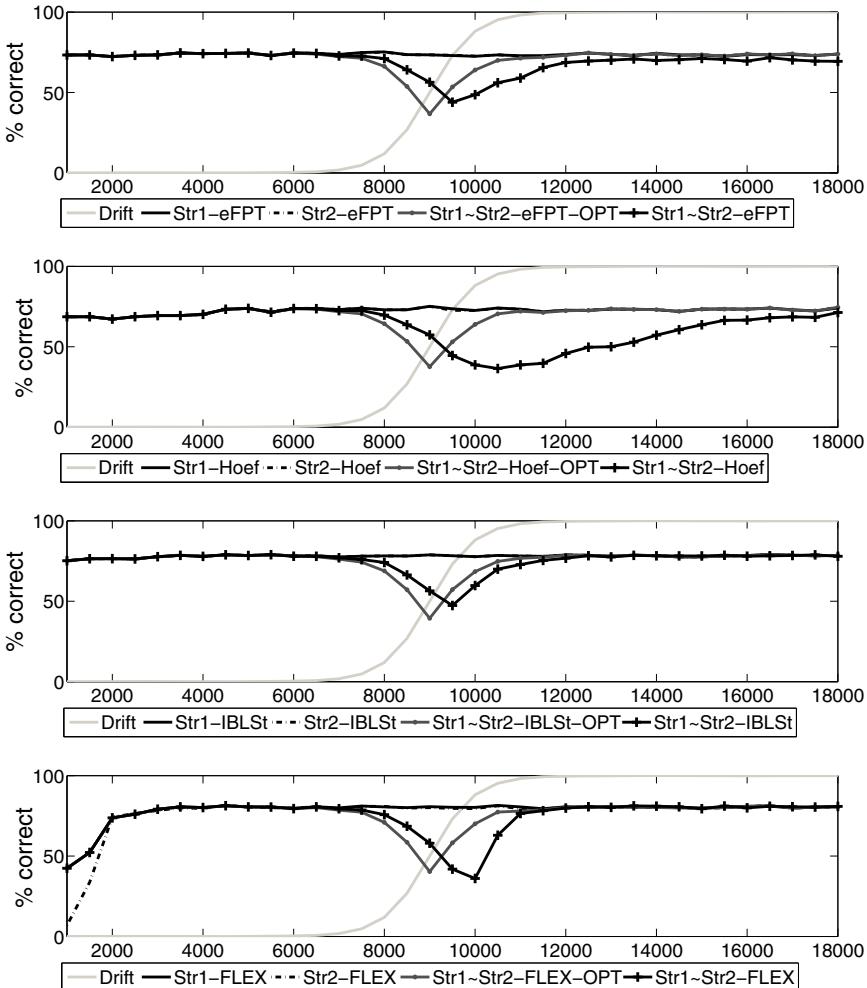


Fig. 2 Performance curves (classification rate) on the weather data. From top to bottom: fuzzy pattern trees, Hoeffding Trees, IBLStreams, FLEXFIS. The sigmoid in light grey indicates the range of the drift. The grey line shows the lower bound on the optimal performance (3).

5 Conclusion

We have introduced *recovery analysis* as a new type of experimental analysis in the context of learning from data streams. The goal of recovery analysis is to provide an idea of a learner's ability to discover a concept drift quickly, and to take appropriate measures to maintain the quality and generalization performance of the model. To demonstrate the usefulness of this type of analysis, we have shown results of an experimental study using a stream of weather data, on which we have compared four different learning algorithms.

In future work, we plan to further refine our approach to recovery analysis, for example by developing numerical measures to quantify specific aspects of a recovery curve (delay of reaction, duration, etc.). Moreover, we plan to use recovery analysis in large empirical studies as a tool for comparing different classes of learning methods with regard to their ability to handle concept drift [13].

Acknowledgements. This work was supported by the German Research Foundation (DFG).

References

1. Ben-David, S., Gehrke, J., Kifer, D.: Detecting change in data streams. In: Proc. VLDB 2004 (2004)
2. Bifet, A., Kirkby, R.: Massive Online Analysis Manual (August 2009)
3. Domingos, P., Hulten, G.: A general framework for mining massive data streams. Journal of Computational and Graphical Statistics 12 (2003)
4. Domingos, P., Hulten, G.: Mining high-speed data streams. In: Proc. KDD 2000, pp. 71–80 (2000)
5. Elwell, R., Polikar, R.: Incremental learning of concept drift in nonstationary environments. IEEE Transactions on Neural Networks 22(10), 1517–1531 (2011)
6. Gaber, M.M., Zaslavsky, A., Krishnaswamy, S.: Mining data streams: A review. ACM SIGMOD Record 34(1) (2005)
7. Gama, J.: A survey on learning from data streams: current and future trends. Progress in Artificial Intelligence 1(1), 45–55 (2012)
8. Gama, J.: Knowledge Discovery from Data Streams. Chapman & Hall/CRC (2010)
9. Gama, J., Gaber, M.M.: Learning from Data Streams. Springer (2007)
10. Lugofer, E.: FLEXFIS: A robust incremental learning approach for evolving Takagi-Sugeno fuzzy models. IEEE Transactions on Fuzzy Systems 16(6), 1393–1410 (2008)
11. Shaker, A., Hüllermeier, E.: IBLStreams: A system for instance-based classification and regression on data streams. Evolving Systems 3(4), 235–249 (2012)
12. Shaker, A., Senge, R., Hüllermeier, E.: Evolving fuzzy pattern trees for binary classification on data streams. Information Sciences 220, 34–45 (2013)
13. Žliobaite, I., Pechenizkiy, M.: Reference framework for handling concept drift: An application perspective. Technical report (2010)
14. Widmer, G., Kubat, M.: Effective learning in dynamic environments by explicit context tracking. In: Brazdil, P.B. (ed.) ECML 1993. LNCS, vol. 667, pp. 227–243. Springer, Heidelberg (1993)

Analysis of Roles and Groups in Blogosphere

Bogdan Gliwa, Anna Zygmunt, and Jarosław Koźlak

Abstract. In the paper different roles of users in social media, taking into consideration their strength of influence and different degrees of cooperativeness, are introduced. Such identified roles are used for the analysis of characteristics of groups of strongly connected entities. The different classes of groups, considering the distribution of roles of users belonging to them, are presented and discussed.

1 Introduction

In recent times an increasing popularity of social media, for example blogosphere, has taken place. It has a more and more increasing importance for marketing forecasting or predicting the popularity of persons and ideas. Dynamic interactions between persons may influence the development of groups of strongly connected entities. Nowadays, applying of the methods of social network analysis is a popular approach for understanding the behaviour of such systems.

The aim of this work was the identification of influential entities and ranges of their influence both on the global and local (in the scope of groups) level. In the analysis, it is assumed that the influence may be built by taking part in discussions regarding personally proposed subjects or by active participation in discussions of subjects proposed by others.

Such analysis may be conducted by recognition of characteristics of existing groups, identification of different kinds of groups by taking into account the presence of the users building their position or taking part in the discussion of other subjects proposed by others. This helps in drawing conclusions concerning the methods of gaining popularity, kinds of influences, their stability and duration.

Bogdan Gliwa · Anna Zygmunt · Jarosław Koźlak
AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: {bgliwa, azygmunt, kozlak}@agh.edu.pl

2 Overview

Roles in blogosphere. There are many definitions of concept of role [12], [3], [13]. In social media, *role* can be treated as a set of characteristics that describe behavior of individuals and their interactions between them within a social context [7]. A lot of studies relate to certain social media and attempt to define their specific roles [9]. For example, an analysis of the basic SNA measures has been used in several studies to define social roles of *starters* and *followers* in blogosphere [6], [8]. *Starters* receive messages mostly from people who are well-connected to each other, and therefore they can be identified by low in-degree, high out-degree and high clustering coefficient in the graph. The distinction between the roles is obtained by combining the difference between the number of in-links and out-links of their blogs.

Groups. Usually it is assumed that the group is a set of vertices which communicate to each other more frequently than with vertices outside the group. Many methods of finding groups have been proposed [2], taking into account the time and its impact on the life cycle of the groups [1] [11]. In [14], [5] algorithm SGCI (Stable Group Changes Identification) was proposed, which we use mainly with CPM (Clique Percolation Method) [10]. The algorithm consists of four main steps: identification of short-lived groups in each separated time interval; identification of group continuation (using modified Jaccard measure), separation of the stable groups (lasting for a certain time interval) and the identification of types of group changes (transition between the states of the stable groups).

3 General Model

The analysed social system is characterised by the following elements: global roles played by given entities in the whole network, groups identified in a network of strongly interacting entities and roles assigned to entities in each group. The analysis of the considered social system is presented and the mentioned elements are calculated in time steps.

Global roles. Global roles present activity and influence of given entities in the scope of the whole social system. They distinguish also the specific features which characterise the behaviour of the entities: post influence, comments influence, local commenting in the context of its own posts or commenting of posts belonging to other users or comments to them. Posts or comments may be considered as influential, which cause a significant response from other bloggers, expressed by numerous comments related to them.

Such a distinction comes from the fact, that there are two kinds of influential entities: (i) the ones who build only their own position and refer only to these users who comment on their posts or comments and (ii) more social ones who take part in other discussions and comment also the posts of other bloggers. Among such influential users it is possible to distinguish also such ones who write both influential

posts and comments (called later *User*) and such who write only the influential posts (called later *Blogger*).

Considering the above criteria, it is possible to distinguish the following important roles: (*Selfish Influential User*) – they build their own position, write influential posts and comments, mainly in the context of their own posts. (*Social Influential User*) – participate also in the contexts of posts written by other users, write influential posts and comments, (*Selfish Influential Blogger*) – they write only influential posts, their comments relate mostly to the context of their own posts, (*Social Influential Blogger*) – they write only influential posts, they are concerned in commenting both their own and other posts.

Among the remaining users it is possible to distinguish those who are no authors of significant posts, but actively comment on posts of others. They are: *Influential Commentator*, writing influential comments and *Standard Commentator*, writing sufficiently numerous, but not necessarily influential comments, and seldom – posts. Users writing very few posts and comments are classified as *Not Active* users.

The rest of the users, who are neither distinguished by the particular activity nor stopped entirely their activity on the portal, have been assigned to the role of *Standard Blogger*.

In the presented social system, especially important is the presence of the influential roles, to which the first four are assigned to.

Stable groups. In each time step the groups are identified using the algorithm for identifying overlapping groups. Then, their structure is compared with the structure of the groups existing in the previous and next step. If a given group may be considered as a continuation of the group from the previous time steps and such continuation exists at least during a given time interval, such a group can be treated as a stable group. The following elements are presented for stable groups.

Roles in groups. For each stable group and each time step, roles of entities in these groups are identified in the similar manner as the identification of roles for the whole social system. The given users in the given time steps may be described by the role played by them globally and by roles played in different groups they belong to. It is also possible to analyse the given users globally from the point of view of the whole considered time period (containing a given number of time steps) and determine the number of roles the user played in this time period globally or in the given groups.

Group characteristics. The groups may be characterised on the basis of percentages of the users having special characteristics, expressed by the sets of roles. They may be as follows:

- users with a high significance (*Selfish Influential User*, *Social Influential User*, *Selfish Influential Blogger*, *Social Influential Blogger*, *Influential Commentator*),
- cooperative users, (*Social Influential User*, *Social Influential Blogger*, *Influential Commentator*),
- users oriented mainly to building their own position in the network (*Selfish Influential User*, *Selfish Influential Blogger*).

Definitions. Let us define *Post Response* (pr_{post}) that it is a number of comments for a given post excluding the author's comments in his own thread. We are using the following notation $c(X, cond)$ that means the number of elements in set X that each element of the set fulfills condition $cond$ and $c(X)$ means the number of all elements in set X . Furthermore, notation $posts_a$ denotes posts of author a , $comments_a$ - comments of user a and $w(cond)$ - expression that returns 1 when the condition $cond$ is satisfied, otherwise this expression returns 0.

Post Influence ($PostInf_a$) describes how influential posts the author a writes, and can be defined as follows:

$$PostInf_a = \sum_{(i,j)} i \cdot c(posts_a, pr \geq j) - \sum_{(k,l)} k \cdot c(posts_a, pr < l) \quad (1)$$

where i,k – weights; j,l – thresholds of influence necessary for assigning a given weight.

Let us define *Comment Response* (cr_a) for author a as a number of comments that refers to comments of author a . Using this term we can formulate *Comment Ratio* ($comRatio_a$) for author a (if the author a wrote at least one comment, otherwise *Comment Ratio* equals 0):

$$comRatio_a = cr_a / comments_a \quad (2)$$

Comment Influence ($ComInf_a$) describes the impact of comments written by author a :

$$ComInf_a = \sum_{(i,j)} i \cdot w(comRatio_a \geq j) - \sum_{(k,l)} k \cdot w(cr_a < l) \quad (3)$$

where i,k – weights; j, l – thresholds.

Comments that author a writes in his own posts are marked as $ownCom_a$. Using this notation we can also define *Comment Ego* ($ComEgo_a$) as:

$$ComEgo_a = ownCom_a / comments_a \quad (4)$$

4 Description of Experiments

Data set. The analysed data set contains data from the portal *salon24*¹. The data set consists of 26 722 users (11 084 of them have their own blog), 285 532 posts and 4 173 457 comments within the period 1.01.2008 - 31.03.2012. The presented results were conducted on half of this dataset - from 4.04.2010 to 31.03.2012. The analysed period was divided into time slots, each lasting 7 days and neighboring slots overlap each other by 4 days. In the examined period there are 182 time slots. In each slot we used the comments model, introduced by us in [4] - the users are nodes and relations

¹ Mainly focused towards politics, www.salon24.pl

between them are built in the following way: from user who wrote the comment to the user who was commented on or if the user whose comment was commented on is not explicitly referenced in the comment (by using @ and name of author of comment) the target of the relation is the author of post.

Role definition. The measures described in the model (section 3) in experiments have the following values:

Post Influence of author a is calculated as follows:

$$\begin{aligned} PostInf_a = & 4c(posts_a, pr \geq A_p) + 2c(posts_a, pr \geq B_p) \\ & + c(posts_a, pr \geq C_p) - c(posts_a, pr < D_p) - 2c(posts_a, pr < E_p) \end{aligned} \quad (5)$$

where A_p, B_p, C_p, D_p and E_p are parameters, describing the strength of the influence assigned to user, when it exceeds the given threshold.

For global roles the following values were used: $A_p=100, B_p=100, C_p=50, D_p=6$ and $E_p=1$. For local roles the parameters depend on the size of group: $A_p=10*\sqrt{\text{size}}$, $B_p=A_p/2, C_p=B_p/2, D_p=0$ and $E_p=1$.

Comment Influence for author a is calculated as follows:

$$\begin{aligned} ComInf_a = & 4w(comRatio_a \geq A_c) + 2w(comRatio_a \geq B_c) \\ & + w(comRatio_a \geq C_c) - w(cr_a < D_c) - 2w(cr_a < E_c) - 4w(cr_a < F_c) \end{aligned} \quad (6)$$

where A_c, B_c, C_c, D_c, E_c and F_c are parameters.

In experiments we used the following values (constant for global and local roles): $A_c=1.25, B_c=1, C_c=0.75$. For global roles we set parameters $D_c=50, E_c=20$ and $F_c=10$, but for local roles the parameters have values as follows (depending from group size): $D_c=2*\sqrt{\text{size}}, E_c=D_c/2, F_c=E_c/2$.

Using terms defined above, we can assign users into one of the following categories:

1. *Influential User (infUser)*: $PostInf > 2$ and $ComInf > 0$
 - a. *Selfish Influential User*: $ComEgo \geq 0.75$
 - b. *Social Influential User*: $ComEgo < 0.75$
2. *Influential Blogger (infBlog)*: $PostInf > 2$ and $ComInf \leq 0$
 - a. *Selfish Influential Blogger*: $ComEgo \geq 0.75$
 - b. *Social Influential Blogger*: $ComEgo < 0.75$
3. *Influential Commentator (infComm)*: $ComInf > 0$ and $PostInf \leq 2$
4. *Standard Commentator (comm)*: $c(\text{comments}) \geq 20$ and $c(\text{posts}) \leq 2$
5. *Not Active (notActive)*: $c(\text{posts}) < 1$ and $c(\text{comments}) < 2$
6. *Standard Blogger (stdBlog)*: User that does not match to any from above roles.

Groups. To extract groups from networks we used CPMd from CFinder² tool, for k equals 5. For group evolution, we took advantage of the SGCI method. Figure 1a

² www.cfinder.org

presents the overall numbers of groups summed up in all time slots and figure 1b shows how the number of groups varies in each time slot. One can see that groups at size 5 outnumber others (overall and in each time slot).

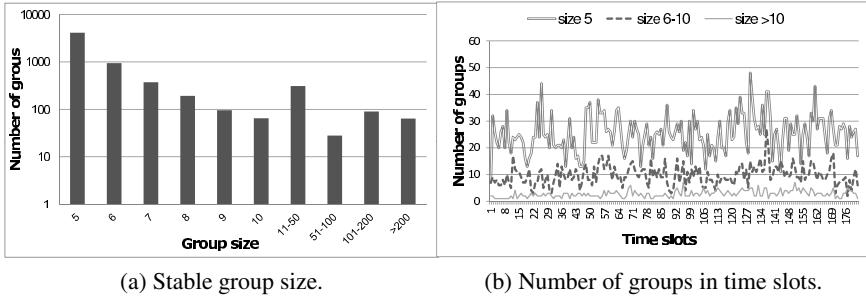
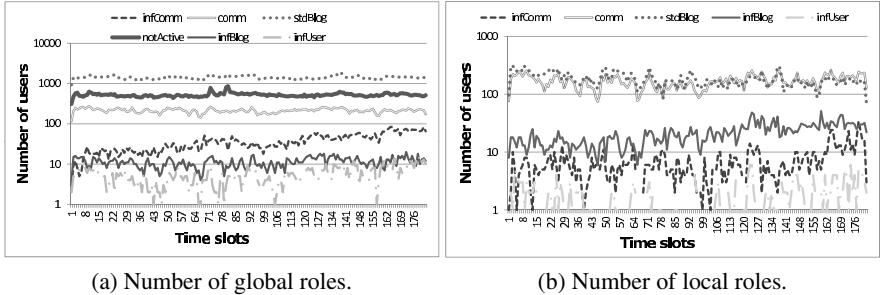


Fig. 1 Summary of stable groups

Roles. In figures 2a and 2b we can see how the roles are evolving during time respectively in whole network (global roles) and in groups (local roles). Both in network and in groups, the *Standard Bloggers* and *Commentators* represent the largest part among all roles. Furthermore, in the whole network there are many inactive users, which is very rare in local roles (a user can be inactive in the given time slot and be a member of a group when in a given time slot where he wrote nothing but he received some comments on what he had written in previous time slots). In whole network among important roles the most numerous are *Influential Commentators*. As we can see in figure 2a the percentage of *Influential Commentators* increases during time - it may have connection with the popularising using '@' in title of comment among users to indicate the response (we showed this trend in paper [4]). In figure 2b we can notice that in groups there are more *Influential Bloggers* than *Influential Commentators* (in contrast to figure 2b) which can suggest that most of the *Influential Commentators* are outside of these groups.

Tab. 1 presents global and local roles for some users placed high in ranking for each global role. We can observe that users who have global role *Influential User Selfish* mainly have local role *Influential Blogger Selfish*. Moreover, *Influential User Selfish* on a global level becomes *Influential Blogger Social* locally. Users that are *Influential Bloggers* globally (both *Social* and *Selfish*) are in the local level also as *Influential Bloggers*. But users playing global roles of *Influential Commentators* in very rare cases have the same role on the local level. It may indicate that influential commentators receive the most responses to their comments from outside their own groups or that these users write a lot of comments outside their group.

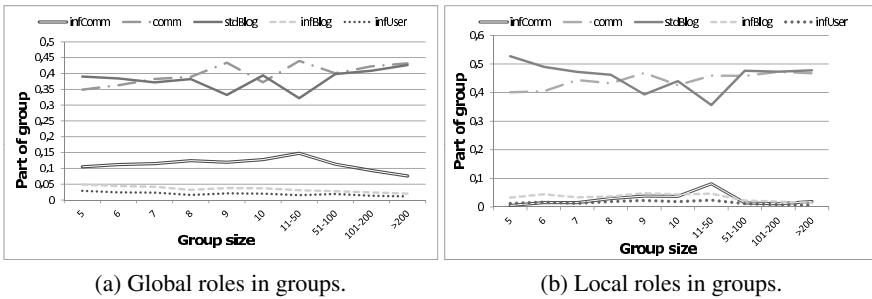
Roles in groups. We considered how large part of the groups is constituted by users with specific roles (role defined on global and local levels). Results can be observed on figures 3a and 3b. The biggest part of the groups constitute not important users (*Standard Bloggers* and *Commentators*), the largest groups (above 100 members) have smaller percentages of important users (both global and local

**Fig. 2** Global and local roles in time slots**Table 1** Roles for selected users (how many times the user had given role, number in brackets means position in ranking)

UserId	Roles	infUserSel	infUserSoc	infBlogSel	infBlogSoc	infComm
2177	global	64(1)	5	14	1	19
	local	23	1	61	4	4
1672	global	41(3)	54(1)	27(15)	3	21
	local	11	19	58	39	14
241	global	41(3)	2	13	0	35
	local	12	1	44	2	12
796	global	13(10)	30(3)	9	17(4)	27
	local	5	6	19	42	4
11	global	1	46(2)	0	38(1)	47(13)
	local	0	1	1	65	0
657	global	0	0	141(1)	1	0
	local	0	0	239	2	0
783	global	3	1	94(2)	1	2
	local	1	0	134	7	0
1991	global	0	0	0	24(2)	0
	local	0	0	0	23	0
7325	global	0	4	0	0	79(1)
	local	0	0	0	9	8
549	global	4	4	7	1	69(2)
	local	0	0	17	41	2

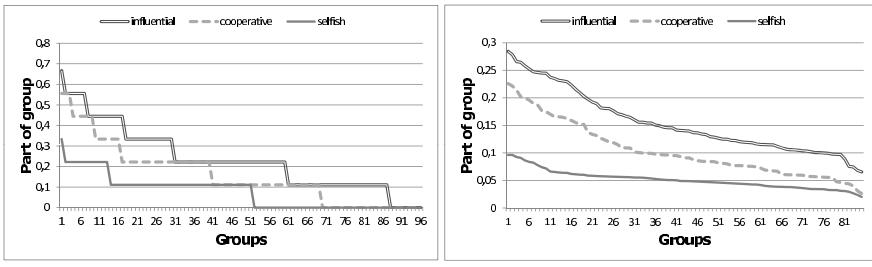
levels) than smaller groups. It may mean that it is easier to play important roles in smaller groups.

Classes of groups. We considered small and large groups separately and look into how many users with given roles are inside them. We decided to take into consideration groups of size 9 as representative of small groups and groups at size between 150 and 250 as representative of large ones. Both cases have similar numbers of groups (there are 96 small groups and 85 large groups). We assigned groups to some classes based on 3 dimensions defined by the percentage of groups that 3 characteristics of role take (*Influential*, *Cooperative* and *Selfish* users). Using distribution of measures presented on figures 4a and 4b we defined threshold levels (for each dimension there are 2 thresholds) showed in tables 2a and 2b. Thresholds were defined by dividing measure distribution into 3 intervals with similar range. For each

**Fig. 3** Global and local roles in groups

dimension the ranges are ordered from 1 to 3 (label 1 concerns range below 1st threshold, label 2 - between 2 threshold, and 3 - above 2nd threshold).

In the results we obtained some classes for each case - tables 3a and 3b. We show the density for classes of small groups and the cohesion for large groups. One can see connections between these measures and classes of groups, for each class we calculated the mean value of density or cohesion for groups in that class. We can notice that more cooperative groups have higher values of density or cohesion for small and large groups (both on a global and local level). In the small groups on the global level, groups with the largest part of selfish roles are less dense.

**Fig. 4** Global roles in small and large groups**Table 2** Thresholds for classes

(a) Small groups.

	Roles	Influential	Cooperative	Selfish
global		0.2,0.4	0.1,0.3	0.1,0.2
local		0.15,0.3	0.1,0.2	0.1,0.15

(b) Large groups.

	Roles	Influential	Cooperative	Selfish
global		0.1,0.2	0.1,0.15	0.05,0.07
local		0.04,0.08	0.02,0.05	0.02,0.04

Table 3 Classes for groups (in rows: global/local)

(a) Classes for small groups.					(b) Classes for large groups.				
Influen.	Cooper.	Self.	Density	Count	Influen.	Coop.	Self.	Cohesion	Count
1	1	1	0.596/0.572	9/34	1	1	1	5.98/6.05	9/20
1	1	2	0.567/0.569	2/19	1	1	2	5.68/5.92	1/9
1	2	1	0.6/0.588	14/6	1	2	1	-6.1	-3
2	1	3	0.555/0.514	5/1	2	1	1	6.21/-	25/-
2	2	1	0.618/-	14/-	2	1	2	6.28/6.07	12/9
2	2	2	0.61/0.604	17/14	2	1	3	5.97/6.8	6/1
2	2	3	0.551/-	3/-	2	2	1	6.57/6.69	8/7
2	3	1	0.559/-	4/-	2	2	2	6.42/6.61	4/15
3	2	3	0.511/0.736	5/1	2	2	3	-6.37	-4
3	3	1	0.635/0.608	4/4	2	3	1	6.44/7.19	1/2
3	3	2	0.646/0.644	8/11	3	2	3	6.44/7.03	1/7
					3	3	1	6.99/6.99	1/1
					3	3	2	7.42/7.42	14/7
					3	3	3	6.76/-	3/-

5 Conclusion

In the paper we presented the research concerning the identification of the important roles described on the basis of different characteristics of the activities in the blogosphere. Configurations of roles in groups and characteristics of such groups were analysed. The roles were considered both on the level of the whole network and for the given groups.

The performed analysis shows that it is possible to distinguish several classes of groups, considering the percentage of significant users belonging to them and their levels of cooperativeness. Groups with different sizes have various behaviour features caused by having a different percentage of influential users. Future works will concern the analysis of the lifespan of groups and identification of the core of groups - their parts that are strongly connected and most stable.

Acknowledgements. The research leading to these results has received funding from the European Community's Seventh Framework Program – Project INDECT (FP7/2007-2013, grant agreement no. 218086).

References

1. Asur, S., Parthasarathy, S., Ucar, D.: An event-based framework for characterizing the evolutionary behavior of interaction graphs. *ACM Trans. Knowl. Discov. Data* 3(4) (2009)
2. Fortunato, S.: Community detection in graphs. *Phys. Rep.*, chap. 486 (2010)
3. Gleave, E., Welser, H., Lento, T., Smith, M.: A conceptual and operational definition of 'social role' in online community. In: 42nd Hawaii International Conference on System Sciences, HICSS 2009, pp. 1–11. IEEE (2009)

4. Gliwa, B., Koźlak, J., Zygmunt, A., Cetnarowicz, K.: Models of social groups in blogosphere based on information about comment addressees and sentiments. In: Aberer, K., Flache, A., Jager, W., Liu, L., Tang, J., Guéret, C. (eds.) SocInfo 2012. LNCS, vol. 7710, pp. 475–488. Springer, Heidelberg (2012)
5. Gliwa, B., Saganowski, S., Zygmunt, A., Bródka, P., Kazienko, P., Kozlak, J.: Identification of group changes in blogosphere. In: ASONAM 2012: IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining: Istanbul, Turkey (2012)
6. Hansen, D.L., Shneiderman, B., Smith, M.: Visualizing threaded conversation networks: Mining message boards and email lists for actionable insights. In: An, A., Lingras, P., Petty, S., Huang, R. (eds.) AMT 2010. LNCS, vol. 6335, pp. 47–62. Springer, Heidelberg (2010)
7. Junquero-Trabado, V., Dominguez-Sal, D.: Building a role search engine for social media. In: Proc. of the 21st Int. Conf. Companion on World Wide Web, WWW 2012 Companion, pp. 1051–1060. ACM, NY (2012)
8. Mathioudakis, M., Koudas, N.: Efficient identification of starters and followers in social media. In: Proc. of the 12th Int. Conf. on Extending Database Technology: Advances in Database Technology, EDBT 2009, pp. 708–719. ACM, NY (2009)
9. Nolker, R., Zhou, L.: Social computing and weighting to identify member roles in online communities. In: Proceedings. The 2005 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 87–93 (2005)
10. Palla, G., Derenyi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 814–818 (2005)
11. Spiliopoulou, M.: Evolution in social networks: A survey. In: Aggarwal, C.C. (ed.) Social Network Data Analytics. Springer (2011)
12. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)
13. Welser, H.T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., Smith, M.: Finding social roles in wikipedia. In: Proc. of the 2011 iConference, iConference 2011, pp. 122–129. ACM, New York (2011)
14. Zygmunt, A., Bródka, P., Kazienko, P., Kozlak, J.: Key person analysis in social communities within the blogosphere. *J. UCS* 18(4), 577–597 (2012)

Knowledge Generalization from Long Sequence of Execution Scenarios

Radosław Z. Ziembinski

Abstract. This paper describes study on generalization of information obtained from context sequential patterns. Problem definition of the context based sequential patterns mining extends the basic sequence structure and introduces heterogeneous attributes describing elements and sequences. Introduction of continuous context information requires knowledge aggregation after the mining. It is a result of many mined patterns that differ only on context attributes values. This paper introduces a new algorithm for the patterns generalization. It aggregates similar patterns and provides compact generalized patterns more readable for humans. Performed experiments shown that the algorithm provides reasonable generalization that accurately represents original knowledge mined from the database.

1 Introduction

Growth of web services provided via Internet creates demand for different integration technologies and new analytical tools. Some of integration technologies rapidly become industry standards like popular Business Process Execution Language. However, accessibility of analytical tools is still a pursued research problem. From this perspective, execution of complex web service can be represented as a short sequence of parametrized calls to business methods. Such an execution log can be assigned to a user and stored in the database. In a more dynamic environment resembling data stream processing, the patterns mining is done directly on execution logs coming immediately from a working web server.

Execution scenarios can be represented as sequences of function calls and parameters passing heterogeneous data. In conjunction with information from users

Radosław Z. Ziembinski

Institute of Computing Science, Poznań University of Technology,
ul. Piotrowo 3A, 60–965 Poznań, Poland

e-mail: radoslaw.ziembinski@cs.put.poznan.pl

profiles they can be stored as context sequences [12] and mined by *ContextMappingImproved* algorithm introduced in [14]. Element of sequence represents a call with the signature encoded in context attributes and items. The mining algorithm delivers a set of context patterns that express frequent sequences of business methods calls. In contradiction to the classic approach proposed in [1] the context algorithm may directly process continuous values of parameters of business methods calls. Thus, it avoids earlier discretization to nominal representation and perturbations related to information loss described in [15].

Table 1 Examples of classic, context and generalized patterns describing shopping

Classic pattern	<(young, man, medium earnings, living loc. A), (loc. A, monday, morning, bread, coffe), (loc. B, friday, morning, milk, bread), (loc. C, potatoes, midday, sunday, sugar, milk)>
Context pattern	<(age: 25, man, income: 2500EUR, living: 51.703N;17.435E), (loc: 51.703N;17.435E, time: 04/06/2012;09:15) (bread, coffe), (loc: 51.772N;18.057E, time: 15/06/2012;08:23) (milk, bread), (loc: 52.309N;17.435E, time: 24/06/2012;10:54) (potatoes, sugar, milk)>
Generalized context pattern	<(age: 23-28, man, income: 1800-2600EUR, living: western Poland), (loc: 51.30-53.45N;16.36-18.27E, time: 02-10/06/2012; 7:50-10:20) (bread, coffe), (loc: 50.23-51.24N;15.23-19.36E, time: 5-14/06/2012; 8:10-09:10) (milk, bread), (loc: 52.24-51.35N;17.36-18.43E, time: 20-28/06/2012; 6:45-12:30) (potatoes, sugar, milk)>

This paper delivers a study of a post-processing step in the patterns mining process. The step produces generalized output from mined context patterns. All preceding steps related to the mining were intentionally omitted in the paper because they are not a subject from the perspective of this study. Currently, there is a lot of published research implementing sliding window design pattern in algorithms exploring different data streams.

A difference between classic, context and generalized patterns makes amount of data involved in the processing and employed data types. Table 1 illustrates example patterns. The classic pattern defined in [1] cannot handle continuous information. It delivers approximation of the ratio scale not preserving its properties. Performed mining is blind on these properties delivering less accurate patterns. On the opposite side are context patterns that give precise values of continuous attributes. However, it is easy to imagine difficulties in harnessing a flurry of possible patterns that differs slightly on values from all sequences in the database. A compromise leads to solution, third kind of patterns derived from second ones. Generalized patterns aggregate exact values from context patterns avoiding discretization before the mining.

The example reveals key assumptions of this research. Firstly, an application of the context approach with the patterns generalization should improve the results accuracy and eliminate a necessity of a prior discretization. The generalization should cut the output set size and keep mined knowledge as much precisely as it is possible. Then, an investigation on relations between patterns values, support and generalization results could give interesting observations about the processing method choice

and configuration. Finally, the research would lead to formulation of a similarity measure for comparing sets of patterns.

The paper begins from a short introduction to related works that precedes definition of the generalization algorithm. Then, the paper describes a method of the similarity evaluation for context patterns sets. Theoretical consideration complements experimental study of the algorithm's results accuracy concluding the paper in final sections.

2 Related Works

The concept of sequential patterns mining algorithms was introduced in paper [1]. It employs comparison of elements encapsulated in non-empty sets of nominal items. For this purpose, algorithm *AprioriAll* uses an inclusion operation. The sequential patterns mining problem has inspired many following researches e.g., see review in [2]. Some authors proposed computationally effective algorithms for mining sequential patterns e.g. PrefixSpan [9] and Spade [13]. Others studied various generalizations (e.g., time constraints, items hierarchy [3] and [10], windows [11], quantities [5]) or applications (e.g., [4], [7], [6]). Context based sequential patterns mining from [12] introduced sets of heterogeneous attributes describing sequences and elements.

The paper delivers following results contributing to the subject. The first one is the context patterns generalization algorithm allowing for compressing of context patterns to more compact and representative set of patterns. Other one is a new measure for evaluation of similarity between context patterns sets. This measure was formulated to compare patterns sets and may be used for their clustering. Similar measures were already proposed e.g., the supporting sequences coverage measure [7] and the items sets based similarity measure [3]. However, none of them exploits context information. A simpler measure considering context information and items sets is proposed in [15].

3 Generalization Algorithm for Context Patterns

The context database D is a snapshot set of context sequences S from data stream T . The sequence consists of ordered sequence of elements $s_i = \{d_i, < e_{i,1}, e_{i,2}, \dots, e_{i,|s_i|} >\}$. Each sequence description is a set of attributes $D = \{A_1, A_2, \dots, A_n\}$. In contradiction to the classic problem definition, context element $e_{i,j} = < i_{i,j}, c_{i,j} >$ has associated a separate set of context attributes $C = \{B_1, B_2, \dots, B_m\}$. Items set residing in element $i_{i,j} \in I$ is left intact, where I is a set of all items. Patterns have the same structure as context sequences. There are two similarity measures and thresholds accompanying contexts $\{\sigma_C, \theta_C\}$ and $\{\sigma_D, \theta_D\}$. Measures accept two contexts instances and returns value in range $[0; 1]$ representing degree of similarity. Value 1 means identity of contexts. It can be stated that contexts c_1 and c_2 come from similar

elements if $\sigma_C(c_1, c_2) \geq \theta_C$ (the same is true for sequence contexts). Hence, element $e_{i,j}$ supports other $e_{k,l}$ if $\sigma_D(d_i, d_k) \geq \theta_D \wedge \sigma_C(c_{i,j}, c_{k,l}) \geq \theta_C \wedge i_{i,j} \subseteq i_{k,l}$. Labels i, k identifies sequences and j, l their elements. Complete problem definition and algorithm called *ContextMappingImproved* can be found in [14]. The shortcomings of the algorithm come from the fact that it mines sparse and large output set of context patterns with significant redundancy at the level of items sets. The context may contain incomplete set of attributes if similarity function is capable to recognize missing values. It allows to implement different sets of parameters from business calls.

Algorithm *CPGen* (Context Patterns Generalization) finds mutually similar patterns with equal sequence of items sets according to following code:

```

function G generalize(P, θD, θC, θS):
P // context patterns
G // generalized context patterns
// algorithm sorts context patterns descending
step.1: P.sort() // according to length and support
// supporting sequence coverage calculates Jaccard coefficient
// for sets of sequences supporting patterns
step.2: M = false
    for i = 1 .. |P| do
        for j = i .. |P| do
            if P[i].equalItemsets(P[j]) and
                (ver. 1)      P[i].similarContexts(P[j], θD, θC) and
                (ver. 2)      P[i].supportingSequenceCoverage(P[j]) ≥ θS then
                    M[i, j] = M[j, i] = true
// find similar patterns on list P
// R contains collections of patterns identifiers
step 3: E = similar(M, P) // mask used in the clustering
        U = {} // temporary stack of similar patterns
        findSimilar(R, M, P, E, U) // find similar patterns
step 4: G = aggregateContexts(R) // provide generalized representation

function bool findSimilar(R, M, P, E, U):
M // binary similarity matrix
P // sorted set of context patterns
E // mask of visited (clustered) patterns
U // stack of patterns identifiers that belongs to a single cluster
// scan for similar context patterns
step 1: for j = 1 .. |P| do
// ignore the same, clustered or short patterns
    if E[j] == true then continue
// finds if pattern is similar to cluster
step 2:    fits = true
            for k in U do
                if M[j, k] == false or P[j].length != P[k].length then
                    fits = false; break
step 3:    if fits == true then
                    E[j] = true
                    U.push(j); stored = findSimilar(R, M, P, E, U); U.pop()
                    if stored == false then E[j] = false
                    flag = true
step 4: if flag == true and U.length > 1 then
            R.add(U.copy); return true
        return false
    
```

Initially, it sorts patterns according to their sizes and support (step 1). This procedure allows to distinguish longer and more supported groups of context patterns in the first iteration. Then similarity matrix M is created. It stores binary information about mutual similarity between each pair of context patterns (step 2). In the first variant the algorithm measures similarity between contexts for given similarity thresholds values and similarity functions. Here, a pair of clustered patterns is similar if their items sets, sequence context and all pairs of context from corresponding elements are similar. This algorithm's variant neglects information about supporting sequences. Instead, the other one considers supporting sequences coverage. It is calculated according to following equation $J_c = (P[i].supSeq \cap P[j].supSeq) / (P[i].supSeq \cup P[j].supSeq)$, where $supSeq$ are supporting sequences. Here, patterns are similar if $J_c \geq \theta_S$. After the algorithm created M , the recursive procedure finds packs of mutually similar patterns according to depth-first strategy (step 3). During aggregation, the mask cell E is set to true for patterns not having similar counterparts (M provides this information). This mask improves the processing by excluding dissimilar context patterns or some with the length less than 2. Algorithm omits such short single element patterns to filter out noise. List R contains sets of patterns identifiers (pointing to clusters). Finally, the algorithm transforms sets identified by R and creates generalized context patterns G (step 4). Transformation procedure is crucial and calculates representation of context values for aggregated context patterns sets.

Step 3 is responsible for finding similar patterns according to following procedure. The recursive procedure traverses the sorted patterns list (step 1). It finds subsequent pattern matching to the forming cluster and still non-clustered (step 2). Indexes of patterns participating in the cluster are kept on stack U (step 3). If the recurrent procedure meets the end of the patterns list then a copy of the collection U joins the list R (step 4).

The cost of the mining is exponential to the number of frequent items in sequences. However, it can be delimited by a sufficiently high value of the minimal support threshold. There are also methods of patterns mining finding k-top patterns allowing to set a demanded deadline for the mining time. The efficiency of the introduced procedure depends on a number of mutually similar patterns. For n_P context patterns the complexity is bound to $O(n_P^2)$ due to costs related to formation of similarity matrix M and searching. It is fast for cases with a high number of mutually similar patterns because they would be collected in initial iterations. The worst case occurs when the similarity matrix has a low number of similar pairs. It can be overcome by removing all patterns without any similar counterpart from the patterns list. If the algorithm applies to the sliding window then the procedure can be repeated iteratively to new patterns in P and existing set of clusters stored in R . New patterns supply existing clusters in R and old obsolete patterns instances would be removed from R . In a result, the cost $O(n_P^2)$ would apply only to new patterns after the window was shifted forward. Number of elements in sequences in the window is delimited by n_P value. In practice, it is even smaller because of the support and similarity thresholds delimit it further.

4 Similarity between Sets of Patterns

The proposed similarity measure compares two sets of patterns. It aggregates a similarity matrix T storing similarity values between pairs of context patterns $P_h[i], P_m[j]$ (from compared sets). This measure aggregates values describing qualitatively different properties using the geometric mean. In a case of elements contexts, substitutability of values describing similarity of the same property has allowed appliance of the arithmetic mean.

Following equation evaluates the similarity between pairs of patterns:

$$T[i][j] = (\sigma_D(P_h[i], P_m[j]) \cdot f_{size}(P_h[i], P_m[j]) \cdot f_{cov}(P_h[i], P_m[j]) \cdot f_{elm}(P_h[i], P_m[j]))^{\frac{1}{4}}$$

$$\begin{aligned} f_{cov}(P_h[i], P_m[j]) &= \frac{|P_h[i].supSeq \cap P_m[j].supSeq|}{|P_h[i].supSeq \cup P_m[j].supSeq|} \\ f_{size}(P_h[i], P_m[j]) &= \frac{\min(P_h[i].length, P_m[j].length)}{\max(P_h[i].length, P_m[j].length)} \\ f_{elm}(P_h[i], P_m[j]) &= \sum_{mv \in V(P_h[i], P_m[j])} \frac{f_{MV}(MV)}{|V(P_h[i], P_m[j])|} \\ f_{MV}(mv) &= \sum_{e_m = \langle i_m, c_m \rangle, e_h = \langle i_h, c_h \rangle \in mv} \frac{\sigma_C(c_m, c_h) \cdot \frac{|i_m \cup i_h|}{|i_m \cap i_h|}}{mv.length} \end{aligned}$$

The measure considers cases where patterns differ in lengths. Then shorter patterns may fit to longer one on many possible ways. All matches are stored in $V(P_h[i], P_m[j])$ where object $mv \in V$ is a list of pairs containing similar elements from both patterns.

Matrix T has complete and precise information about similarity between two sets of patterns. It aggregates to a single value according to following equation:

$$f_{avg}(P_h, P_m) = \sum_{i=1..|P_h|} \sum_{j=1..|P_m|} \frac{T[i][j]}{|P_h| \cdot |P_m|}$$

The function evaluates compactness of sets and their similarity at the same time. It returns 1 for two sets containing the same pattern. For internally redundant sets of patterns the matrix became bigger thus the result would be smaller.

5 Experimental Evaluation

An experimental evaluation verified accuracy of the mining and generalization on artificial data sets. The generator creates a specified number of sequences of equal

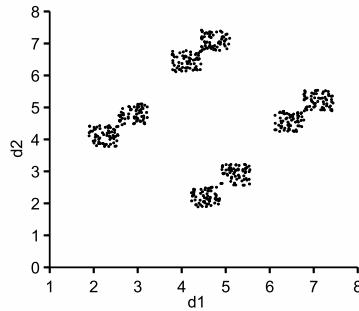


Fig. 1 Example of sequence contexts distribution in the generated database (512 contexts, 2 dimensions, 8 clusters)

lengths that may represent business methods calls with continuous parameters visible in the stream window. It inserts to sequences a specified number of context patterns obtained from templates. Templates have unique sets of items and context values describing sequences and elements. These values undergo random rotations and distortions to produce a pattern from the template inserted to the sequence. However, the distortion is sufficiently small to prevent against contexts values overlapping as illustrated on Fig. 1.

The mining and the generalization used a following formula as the similarity measure:

$$\sigma(c_1, c_2) = \sum_{i=1..|C|: b_{1,i} \in c_1, b_{2,i} \in c_2} \frac{1}{|C|} \cdot \begin{cases} 0.0 & \text{if } |b_{1,i} - b_{2,i}| \geq 1.0 \\ 1.0 & \text{if } b_{1,i} = b_{2,i} \\ |b_{1,i} - b_{2,i}| & \text{if } 0.0 < |b_{1,i} - b_{2,i}| < 1.0 \end{cases}$$

Configuration sets up the window as containing 512 sequences populated from 8 templates (64 sequences each, support 0.125). Each generated element contains 4 items including ones from hidden patterns. Computations involved 16 repetitions for databases generated at the same configuration. Hence, presented results are averages.

The first experiment focused on the accuracy comparison. Considered were context patterns A , patterns generalized using context similarity thresholds B and using the coverage measure C . At the experiment, similarity thresholds had 2 different values at the mining $\theta_C = \theta_D = 0.50$ and $\theta_C = \theta_D = 0.60$. Labels A,B,B/1 and A,B,C/2 denote them in Table 2. Cases B used similarity thresholds values for the generalization, too. For the third case labeled C/1 and C/2 the coverage threshold θ_S was 0.50. Results containing f_{avg} and averaged patterns lengths are in Table 2. They suggest that the most accurate method was one with the generalization adopting the supporting sequence coverage. It also provided smaller sets of generalized context patterns (a few generalized patterns from hundreds context patterns). For $\theta_{sup} = 0.12$ the number of patterns for the method C/1 was very close to the original number of templates populating the database. In a case of 128 sequences the

Table 2 Evaluation of patterns before (A) and after (B,C) the generalization

θ_{sup}	f_{avg}						Average patterns numbers					
	A/1	A/2	B/1	B/2	C/1	C/2	A/1	A/2	B/1	B/2	C/1	C/2
128 sequences, pattern length 4, 4 items, 2 dimensional contexts												
0.220	0.008	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000
0.200	0.058	0.000	0.000	0.000	0.000	0.000	0.625	0.000	0.000	0.000	0.000	0.000
0.180	0.129	0.000	0.000	0.000	0.000	0.000	3.750	0.000	0.000	0.000	0.000	0.000
0.160	0.125	0.000	0.008	0.000	0.010	0.000	14.938	0.000	0.063	0.000	0.063	0.000
0.140	0.124	0.068	0.111	0.000	0.128	0.000	45.500	1.500	1.813	0.000	1.750	0.000
0.120	0.120	0.117	0.123	0.000	0.147	0.000	81.438	8.000	3.813	0.000	3.000	0.000
0.110	0.116	0.114	0.118	0.009	0.146	0.010	111.813	16.063	6.250	0.063	5.125	0.063
0.100	0.108	0.105	0.115	0.060	0.154	0.075	180.938	50.875	11.375	0.875	10.000	0.875
512 sequences, pattern length 4, 4 items, 2 dimensional contexts												
0.220	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.200	0.008	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000
0.180	0.120	0.000	0.000	0.000	0.000	0.000	2.500	0.000	0.000	0.000	0.000	0.000
0.160	0.125	0.000	0.000	0.000	0.000	0.000	40.063	0.000	0.000	0.000	0.000	0.000
0.140	0.122	0.008	0.043	0.000	0.040	0.000	140.125	0.125	0.438	0.000	0.375	0.000
0.120	0.120	0.075	0.131	0.000	0.157	0.000	308.250	2.625	9.375	0.000	8.500	0.000
0.110	0.118	0.118	0.123	0.000	0.156	0.000	449.375	15.750	16.813	0.000	15.188	0.000
0.100	0.112	0.113	0.120	0.000	0.159	0.000	671.688	60.438	18.688	0.000	18.063	0.000

Table 3 The generalization accuracy for different values of the coverage threshold

θ_{sup}	f_{avg}						Average patterns numbers					
	A	$\theta_S = 0.2$	$\theta_S = 0.4$	$\theta_S = 0.6$	A	$\theta_S = 0.2$	$\theta_S = 0.4$	$\theta_S = 0.6$	A	$\theta_S = 0.2$	$\theta_S = 0.4$	$\theta_S = 0.6$
512 sequences, pattern length 4, 4 items, 2 dimensional contexts												
0.20	0.0083	0.0000	0.0000	0.0000	0.0625	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.18	0.0883	0.0000	0.0000	0.0000	2.5625	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.16	0.1250	0.0000	0.0000	0.0000	40.5625	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.14	0.1225	0.0503	0.0503	0.0394	134.2500	0.6875	0.6875	0.4375	0.6875	0.6875	0.4375	0.4375
0.12	0.1201	0.1576	0.1571	0.1585	306.2500	8.9375	8.8750	6.0625	8.9375	8.8750	6.0625	6.0625
0.11	0.1178	0.1583	0.1576	0.1570	436.6250	16.3750	15.5000	11.4375	16.3750	15.5000	11.4375	11.4375
0.10	0.1118	0.1589	0.1580	0.1573	663.8125	19.9375	19.0625	17.0000	19.9375	19.0625	17.0000	17.0000

experiment provided less accurate results. A low number of context patterns that took part in the generalization resulted in worse accuracy. This suggests that the window size must be appropriately large. For unmodified set of context patterns f_{avg} is low even if it was the exact mined knowledge. That remains true but template patterns were randomly disturbed producing nontrivial redundancy for simple context values enumeration. Therefore output set of patterns become large and f_{avg} measure value become low in accordance to its properties (it punishes redundant sets).

The second experiment was conducted to measure performance complexity of the generalization algorithm (C) for different coverage threshold values θ_S . Table 3

presents results. The similarity thresholds were $\theta_D = \theta_C = 0.50$ (were used only at the mining). The difference between settings is not clear for the f_{tavg} values. However, numbers of patterns in the mined and generalized sets differ significantly. Higher value caused that the generalized patterns set became more compact. Such setting causes more accurate clustering of context patterns on the base of their supporting sequences. It may lead to more equal separation of patterns among groups that undergo generalization. Thus, it better eliminates patterns dissimilar to others from the output set.

6 Conclusions

Data mining in sets containing complex, heterogeneous objects is the important research topic with great scope of applicability. Example application includes analysis of execution logs e.g., from BPEL script finding compact patterns describing web services behavior. The mining in context database may involve enumeration of a significant number of patterns. Therefore, the algorithm for the generalization of context patterns is proposed and verified in this paper. It can also produce a compact representation transformable to chain-graph representing parallel and serial dependencies between frequent elements from patterns.

The paper introduces a new accuracy measure capable to compare two sets of context patterns. It is useful for context based clustering employed in the patterns generalization algorithm. Conducted experiments shown that similarity measure based on the supporting sequences coverage proved to be more reliable than one relying only on attributes values similarity. It provided more compact sets of generalized patterns and delivers the most accurate result.

This paper is a result of the project financed by National Science Centre in Poland grant no. DEC-2011/03/D/ST6/01621.

References

1. Agrawal, R., Sirkant, R.: Mining sequential patterns. In: Proc. of the Eleventh International Conference on Data Engineering. IEEE (1995)
2. Dong, G., Pei, J.: Sequence Data Mining. In: Advances in Database Systems. Kluwer (2007)
3. Garofalakis, M.N., Rastogi, R., Shim, K.: SPIRIT: Sequential Pattern Mining with Regular Expression Constraints. In: Proc. of the 25th Int. Conf. on Very Large Data Bases. Morgan Kaufmann (1999)
4. Guralnik, V., Karypis, G.: A Scalable Algorithm for Clustering Sequential Data. In: Proc. of the 2001 IEEE Int. Conf. on Data Mining. IEEE (2001)
5. Kim, C., Lim, J.-H., Ng, R.T., Shim, K.: SQUIRE: Sequential pattern mining with quantities. J. Syst. Softw. (2007)

6. Lesh, N., Zaki, M.J., Ogihara, M.: Mining features for sequence classification. In: Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM (1999)
7. Morzy, T., Wojciechowski, M., Zakrzewicz, M.: Pattern-oriented hierarchical clustering. In: Eder, J., Rozman, I., Welzer, T. (eds.) ADBIS 1999. LNCS, vol. 1691, pp. 179–190. Springer, Heidelberg (1999)
8. Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., Hsu, M.-C.: PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth. In: Proc. of the 17th International Conference on Data Engineering. IEEE (2001)
9. Pinto, H., Han, J., Pei, J., Wang, K., Chen, Q., Dayal, U.: Multi-Dimensional Sequential Pattern Mining. In: Proc. 10th Conf. on Information and Knowledge Management. ACM (2001)
10. Plantevit, M., Laurent, A., Laurent, D., Teisseire, M., Choong, Y.W.: Mining multidimensional and multilevel sequential patterns. ACM Trans. Knowl. Discov. Data (2010)
11. Srikanth, R., Agrawal, R.: Mining Sequential Patterns: Generalizations and Performance Improvements. In: Apers, P.M.G., Bouzeghoub, M., Gardarin, G. (eds.) EDBT 1996. LNCS, vol. 1057, pp. 1–17. Springer, Heidelberg (1996)
12. Stefanowski, J., Ziembinski, R.: Mining context based sequential patterns. In: Szczepaniak, P.S., Kacprzyk, J., Niewiadomski, A. (eds.) AWIC 2005. LNCS (LNAI), vol. 3528, pp. 401–407. Springer, Heidelberg (2005)
13. Zaki, M.J.: SPADE: An Efficient Algorithm for Mining Frequent Sequences. Mach. Learn. (2001)
14. Ziembinski, R.Z.: Algorithms for Context Based Sequential Pattern Mining. Fundamenta Informaticae (2007)
15. Ziembinski, R.Z.: Accuracy of generalized context patterns in the context based sequential patterns mining. Control and Cybernetics, Polish Academy of Science (2011)

Incremental Learning and Forgetting in One-Class Classifiers for Data Streams

Bartosz Krawczyk and Michał Woźniak

Abstract. One-class classification and novelty detection is an important task in processing data streams. Standard algorithms used for this task cannot efficiently handle the changing environment to which they are applied. In this paper we present a modification of Weighted One-Class Support Vector Machine that is able to swiftly adapt to changes in data. This was achieved by extending this classifier by adding incremental learning and forgetting procedures. Both addition of new incoming data and removal of outdated objects is carried out on the basis of modifying weights assigned to each observation. We propose two methods for assigning weights to incoming data and two methods for removing the old objects. These approaches work gradually, therefore preserving useful characteristic of the examined dataset from previous iterations. Our approach was tested on two real-life dynamic datasets and the results prove the quality of our proposal.

Keywords: machine learning, one-class classification, data streams, concept drift, incremental learning, forgetting.

1 Introduction

One-class classification (OCC) is one of the most challenging areas of machine learning. It is assumed that during the classifier training stage we have at our disposal objects coming from a single class distribution, referred to as the target concept. During the exploitation phase of such a classifier there may appear new objects, that were unseen during the classifier building step. They are known as outliers. Therefore one-class classification aims at deriving a classification boundary that may separate the known target objects from possible outliers that may appear.

Bartosz Krawczyk · Michał Woźniak

Department of Systems and Computer Networks, Wrocław University of Technology,
Wyb. Wyspińskiego 27, 50-370 Wrocław, Poland

e-mail: {bartosz.krawczyk, michał.wozniak}@pwr.wroc.pl

No assumptions about the nature of outliers should be made. The term single-class classification originates from [10], but also outlier detection or novelty detection [4] are used to name this field of study.

This is a difficult task that leads to many open problems. How the boundary should be tuned - if it is too general many unwanted outliers would be accepted, if it is to fitted to the training set then a strong overfitting may occur. Therefore it is risky to rely only on a single given model and in recent years there have been several successful attempts on how to improve the quality of one-class recognition systems [11, 12].

In the beginning the data streams originated in the financial markets. Today, the data streams are to be found everywhere - in the Internet, monitoring systems, sensor networks and other domain [9]. The data stream is very different from the traditional static data. It is an infinite amount of data that continuously arrives and the processing time is of a crucial value. Mining data streams poses many new challenges [1].

Most of the existing work on OCC has not explicitly dealt with the changing nature of the input data [14]. They are based on an underlying assumption that the training data set does not contain any uncertainty information and properly represents the examined concept. However, data in many real-world applications change their nature over time - which is a problem frequent for data streams [8]. For example, in environmental monitoring applications data may change according to the examined conditions and what once was considered an outlier may in near future become a representative of the target concept. This kind of dynamic information, typically ignored in most of the existing one-class learning methods, is critically important for capturing the full picture of the examined phenomenon. Therefore there is a need for introducing novel, adaptive techniques for dealing with non-stationary data sets [7].

In this paper we present a novel adaptive Weighted One-Class Support Vector Machine that is able to change itself according to the nature of received data streams. We propose to use the principles of incremental learning and forgetting to allow the changes in the shape of the decision boundary for the new chunks of data. The learning and forgetting in data streams is realized by modifying weights assigned to objects - we propose how to calculate weighs for new incoming objects in order to use their information to change the classifier and how to forget the old objects to prevent the overfitting of the classifier and uncontrolled increase in the volume of the dataset.

2 Weighted One-Class Support Vector Machine

One-class classification aims at distinguishing between the available objects coming from the target distribution ω_T and unknown outliers ω_O , that are unavailable during the classifier training step but may appear in the process of classifier exploitation. One-Class Support Vector Machine (OCSVM) [16] achieves this goal by computing

a closed boundary in a form of a hypersphere enclosing all the objects from ω_T . During the exploitation phase a decision made about the new object is based upon checking whether it falls inside the hypersphere. If so, the new object is labeled as one belonging to ω_T . Otherwise it belongs to ω_O .

The center a and a radius R are the two parameters that are sufficient for describing such a decision hypersphere. To have a low acceptance of the possible outliers the volume of this d -dimensional hypersphere, which is proportional to R^d , should be minimized in such a way that tightly encompasses all available objects from ω_T . The minimization of R^d implies minimization with respect to R^2 . Following this the minimization functional may be formulated as follows:

$$\Theta(a, R) = R^2, \quad (1)$$

with respect to the constraint:

$$\forall_{1 \leq i \leq N} : \|x_i - a\|^2 \leq R^2, \quad (2)$$

where x_i are objects from ω_T , and, N stands for the quantity of training objects. Additionally to allow the fact that there may have been some outliers in the training set and to increase the robustness of the trained classifier some objects with distance to a greater than R are allowed in the training set, but associated with an additional penalty factor. This is done identically as in a standard SVM by the introduction of slack variables ξ_i .

This concept can be further extended to a Weighted One-Class Support Vector Machine (WOCSVM) [3] by the introduction of weights w_i that allows for an association of an importance measure to each of the training objects. This forces slack variables ξ_i to be additionally controlled by w_i . If with object x_i there is associated a small weight w_i then the corresponding slack variable ξ_i indicates a small penalty. In effect, the corresponding slack variable will be larger, allowing x_i to lie further from the center a of the hypersphere. This reduces an impact of x_i on the shape of a decision boundary of WOCSVM.

By using the above mentioned ideas we can modify the minimization functional:

$$\Theta(a, R) = R^2 + C \sum_{i=1}^N w_i \xi_i, \quad (3)$$

with the modified constraints that almost all objects are within the hypersphere:

$$\forall_{1 \leq i \leq N} : \|x_i - a\|^2 \leq R^2 + \xi_i, \quad (4)$$

where $\xi_i \geq 0$, $0 \leq w_i \leq 1$. Here C stands for a parameter that controls the optimization process - the larger C , the less outliers are allowed with the increase of the volume of the hypersphere.

For establishing weights we may use techniques dedicated to a weighted multi-class support vector machines [5]. In this paper we propose to use a following formula:

$$w_i = \frac{|x_i - x_{mean}|}{R + \delta}, \quad (5)$$

where $\delta > 0$ is used to prevent the case of $w_i = 0$.

3 Incremental Learning and Forgetting in One-Class Classification

We assume that the classified data stream is given in a form of data chunks. At the beginning we have at our disposal an initial dataset $\mathcal{D}\mathcal{S}_0$ that allows for training the first phase of classifier. Then with each k -th iteration we receive an additional chunk of data labeled as $\mathcal{D}\mathcal{S}_k$. We assume that there is a possibility of concept drift presence in the incoming chunks of data. Therefore it would be valuable to adjust our one-class classifier to the changes in the nature of data.

In case when we will be using a WOCSVM trained on $\mathcal{D}\mathcal{S}_0$ for all new incoming data we will notice a significant drop in performance - and after few new chunks of data it is possible that the performance of our model will drop below the quality of a random classifier. To prevent this from happening we propose to adapt incrementally the one-class classifier with the new incoming data to deal with the presence of concept drift and allow for a more efficient novelty detection in data streams.

We propose to apply the classifier adaptation in a changing environment via modification of weights assigned to objects from the dataset. We propose the incremental learning procedure, meaning that the dataset $\mathcal{D}\mathcal{S}$ will consist of all available chunks of data at given k -th moment.

3.1 Incremental Learning

We propose to extend the WOCSVM concept by adding an incremental learning principle to it [15]. We use passive incremental learning. In this method new data are added without considering the importance of data. Namely, all new data are added to the training dataset.

As we associate the process of incremental learning with weights assigned to objects we propose to modify the original decision boundary by two methods for calculating weights for objects coming from a new data chunk $\mathcal{D}\mathcal{S}_k$:

- assigning weights to objects from $\mathcal{D}\mathcal{S}_k$ according to eq. (5). This is motivated by the fact that in the incoming data chunk not all objects should have the same impact on the shape of a new decision boundary - there may be outliers or redundant objects present.
- assigning highest weights to objects coming from the new data chunk:

$$\forall_{x_i \in \mathcal{D}\mathcal{S}_k} : w_i = 1. \quad (6)$$

This is motivated by the fact that in the presence of the concept drift objects from a new chunk of data represent the current state of the analyzed dynamic environment and therefore should have a top priority in forming the new decision boundary.

3.2 Incremental Forgetting

If we apply only the incremental learning principle to the WOCSVM, the decision boundary will become more and more complex with each additional chunk of data, that enlarges our data set. This leads to a poor generalization ability in general. This can be avoided by forgetting unnecessary, outdated data [6]. It seems natural that the degree of importance of data reduces as the time passes. We propose to incorporate this concept into our concept by further expanding the WOCSVM with the incremental forgetting principle.

The simplest way is a removal of objects coming from the previous (or oldest iteration). Yet in this way we discard all the information they carried - while they still may have some valuable influence on the classification boundary (e.g., in case of a gradual concept drift where the changes of the data distribution are not rapid). A method that allows for a gradual decrease of the object influence over time seems a far more attractive idea.

Identically as in incremental learning we modify the weights to change the influence of the data on the shape of the decision boundary. In this case we propose to reduce the weights of objects from previous chunks of data in each iteration. After some given number of iterations weights assigned to them should be equal to 0 - meaning that the examined object has no longer any influence on the WOCSVM and can be safely removed from the dataset.

We propose two methods for calculating new weights for objects coming from previous iterations:

- gradual decrease of weights with the respect to their initial importance - here we introduce a denomination factor τ that is a user-specified value by which the weights will be decreased in each iteration:

$$w_i^{k+1} = w_i^k - \tau. \quad (7)$$

This is motivated by the fact that if an object had initially assigned a higher weight it had a bigger importance for the classifier. As such these objects can be valuable for a longer period of time than objects with initial low weights - in this approach their weights will sooner approach the 0 value and they will be removed in a fewer iterations than objects with high initial value of weights.

- aligned decrease of weights without considering their initial importance - here we introduce a time factor κ that is a user-specified value standing for a number of iterations after which the object should be removed:

$$w_i^{k+1} = w_i^k - (w_i^a / \kappa), \quad (8)$$

where w_i^a stands for the initial value of the weight assigned to i -th object. As we can see the weights of objects are reduced with each iteration till they are equal to 0 (and removed from $\mathcal{D}S$) - the main difference is that this method does not consider the initial importance of data. This means that all the objects from k -th data chunk will be removed in the same moment, after κ iterations. This is motivated by the fact that changes in the dynamic environment can be unpredictable and quickly move from the original distribution - therefore data from previous steps may quickly lose its importance.

4 Experimental Investigations

4.1 Set-Up

The aims of the experiment were to assess if embedding an incremental learning and forgetting in a one-class classifier by modifying weights allows to handle the changing data streams and to compare the effectiveness of different learning and forgetting techniques introduced in this paper.

We have used two real-life datasets for the experiment:

- ECUE spam database [18] - the dataset is a collection of spam and legitimate consists of emails received by one individual. Apart from the initial training dataset there are available 12 data chunks with the presence of a gradual concept drift, consisting of messages collected over one year - single month collection in a single data chunk.
- Ozon level detection database - this data describes local ozone peak prediction, that is based on eight hours measurement with the presence of a gradual concept drift.

Each dataset was prepared to consist of 5 parts - one for initial training and five data chunks for testing the incremental learning and forgetting procedures.

For the experiment a WOCSVM with a RBF kernel and Canberra distance metric [13] is used as a base classifier. We have examined the performance of five different WOCSVM models for data stream classification:

1. L_0F_0 - an WOCSVM trained without modifying the values of weights. Here objects from new dataset are added to the training set and the forgetting was implemented as a passive forgetting - objects are removed from the dataset after k iterations. This is a baseline model for further comparison.
2. L_1F_1 - a WCOSVM with the incremental learning by assigning weights to new objects according to eq. (5) and with forgetting by the gradual decrease of weights.
3. L_2F_1 - a WCOSVM with the incremental learning by assigning highest weights to new objects and with forgetting by the gradual decrease of weights.

4. L_1F_2 - a WCOSVM with the incremental learning by assigning weights to new objects according to eq. (5) and with forgetting by the aligned decrease of weights without considering their initial importance.
5. L_2F_2 - a WCOSVM with the incremental learning by assigning highest weights to new objects and with forgetting by the aligned decrease of weights without considering their initial importance.

The value of τ parameter was set to 0.2 and κ to 2. These values were derived using a grid-search procedure, as they offered the best performance.

The combined 5x2 cv F test [2] was carried out to asses the statistical significance of obtained results.

All experiments were carried out in the R environment [17].

4.2 Results

Results for the ECUE dataset are given in Tab. 1, while for the Ozon dataset in Tab. 2.

Table 1 Results of the experiment with the respect to the accuracy [%] and statistical significance for ECUE dataset. Small numbers under each method stands for the indexes of models which the considered one outperforms (in a statistically significant way).

Method	$\mathcal{D}\mathcal{S}_0$	$\mathcal{D}\mathcal{S}_1$	$\mathcal{D}\mathcal{S}_2$	$\mathcal{D}\mathcal{S}_3$	$\mathcal{D}\mathcal{S}_4$
$L_0F_0^1$	84.56	76.34	73.25	73.03	72.44
$L_1F_1^2$	—	—	—	—	—
$L_1F_1^2$	84.56	80.34	79.65	79.22	77.43
$L_2F_1^3$	—	1,4	1,4	1,4,5	1
$L_2F_1^3$	84.56	82.11	81.03	80.04	79.05
$L_1F_2^4$	—	ALL	ALL	1,4,5	1,2
$L_1F_2^4$	84.56	78.96	78.23	78.05	78.20
$L_1F_2^5$	—	1	1	1	1,2
$L_1F_2^5$	84.56	81.54	80.15	78.71	78.55
	—	1,2,4	1,4	1	1,2

4.3 Results Discussion

From the experimental results one may see that proposed methods for incremental learning and forgetting for WOCSVM are a valuable tool for dealing with the changing environment in data streams. The standard OCSVM with simple method for adapting to new data was unable to deal with gradual concept drift present in the examined datasets.

Four tested models based on two learning and two forgetting procedures outperformed this baseline solution on all data. Modifying the weights assigned to objects instead of simply adding them and removing allowed for a smoother change of

Table 2 Results of the experiment with the respect to the accuracy [%] and statistical significance for Ozone dataset. Small numbers under each method stands for the indexes of models which the considered one outperforms (in a statistically significant way).

Method	$\mathcal{D}\mathcal{S}_0$	$\mathcal{D}\mathcal{S}_1$	$\mathcal{D}\mathcal{S}_2$	$\mathcal{D}\mathcal{S}_3$	$\mathcal{D}\mathcal{S}_4$
$L_0F_0^1$	87.44	84.05	82.90	79.45	76.48
$L_1F_1^2$	—	—	—	—	—
$L_1F_1^2$	87.44	85.92	84.11	81.68	80.04
$L_2F_1^3$	—	1	1	1,5	1
$L_2F_1^3$	87.44	86.25	85.55	82.65	83.11
$L_1F_2^4$	—	ALL	ALL	1,2,5	ALL
$L_1F_2^4$	87.44	85.22	84.86	82.70	80.15
$L_1F_2^5$	—	1	1,2	1,2,5	1
$L_1F_2^5$	87.44	85.80	84.89	80.97	81.98
	—	1	1,2	1	1,2,4

model and introduced elasticity into OCC stream data classification. Weight manipulation lead to a situation in which the data from previous chunks had neither too big or to small influence on the shape of the decision boundary. Therefore this approach preserved the valuable influence of the older data, while adapting to changes in the incoming objects.

Out of four tested combination the most promising results were returned by the model based on the incremental learning by assigning highest weights to new objects and the forgetting by the gradual decrease of weights. In most cases it statistically outperformed all other models. This means that during the concept drift the new objects should have the biggest influence on the shape of the new boundary. As for the forgetting procedure the experiments favored the approach in which relevant objects (i.e., with high initial weights) influence the decision boundary for larger number of iterations than those with low initial weights.

Interestingly there was no stable trend in differences between all other methods, which leads to a conclusion that their performance is data set-related and needs further experimental analysis.

The model response to the presence of the concept drift may display itself by the change of the center or/and radius of the hypersphere. In both experiments the proposed methods responded by a continuous shift of the sphere center towards the direction of the drift. In comparison the radius of the hypersphere was subject to much lower variance.

5 Conclusions

One-class classification and novelty detection in data streams is a promising research direction. There is an ongoing need for introducing classifier models that

can adapt to changing environment. In this paper we have introduced a modified Weighted One-Class Support Vector Machine, augmented with the principles of incremental learning and forgetting. These techniques allowed to adapt the shape of the decision boundary to changes in the stream of data.

We proposed to adapt WOCSVM by modifying weights that are assigned to objects in the set. We have applied the incremental learning by two approaches based on calculating weights for incoming objects. Incremental forgetting was applied to avoid constructing too complex model and to reduce the volume of the dataset, which is an important problem in distributed applications. Forgetting was applied as decreasing the weights assigned to objects over time, to a point in which they no longer influence the WOCSVM and may be removed from the data set.

In our future works we would like to test our approach on different types of concept drift, improve the incremental learning and forgetting procedure by making it selective (i.e., choosing only some part of the data to add to the set or remove from it) and combining our incremental WOCSVM in ensembles.

Acknowledgements. This work is supported by the Polish National Science Center under a grant N N519 650440 for the period 2011-2014.

References

1. Aggarwal, C.C., Han, J., Wang, J., Yu, P.S.: On demand classification of data streams. In: KDD 2004 Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 503–508 (2004)
2. Alpaydin, E.: Combined 5×2 cv f test for comparing supervised classification learning algorithms. *Neural Computation* 11(8), 1885–1892 (1999)
3. Bicego, M., Figueiredo, M.A.T.: Soft clustering using weighted one-class support vector machines. *Pattern Recognition* 42(1), 27–32 (2009)
4. Bishop, C.M.: Novelty detection and neural network validation. *IEE Proceedings: Vision, Image and Signal Processing* 141(4), 217–222 (1994)
5. Cyganek, B.: One-class support vector ensembles for image segmentation and classification. *Journal of Mathematical Imaging and Vision* 42(2-3), 103–117 (2012)
6. Gent, I.P., Miguel, I., Moore, N.C.A.: An empirical study of learning and forgetting constraints. *AI Communications* 25(2), 191–208 (2012)
7. Gomez-Verdejo, V., Arenas-Garcia, J., Lazaro-Gredilla, M., Navia-Vazquez, A.: Adaptive one-class support vector machine. *IEEE Transactions on Signal Processing* 59(6), 2975–2981 (2011)
8. Hashemi, S., Yang, Y., Mirzamomen, Z., Kangavari, M.: Adapted one-versus-all decision trees for data stream classification. *IEEE Transactions on Knowledge and Data Engineering* 21(5), 624–637 (2009)
9. Hulten, G., Spencer, L., Domingos, P.: Mining time-changing data streams. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 97–106 (2001)
10. Koch, M.W., Moya, M.M., Hostetler, L.D., Fogler, R.J.: Cueing, feature discovery, and one-class learning for synthetic aperture radar automatic target recognition. *Neural Networks* 8(7-8), 1081–1102 (1995)

11. Krawczyk, B.: Diversity in ensembles for one-class classification. In: Pechenizkiy, M., Wojciechowski, M. (eds.) *New Trends in Databases & Inform.* AISC, vol. 185, pp. 119–129. Springer, Heidelberg (2012)
12. Krawczyk, B., Woźniak, M.: Combining diverse one-class classifiers. In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B. (eds.) *HAIS 2012, Part II.* LNCS, vol. 7209, pp. 590–601. Springer, Heidelberg (2012)
13. Krawczyk, B., Woźniak, M.: Experiments on distance measures for combining one-class classifiers. In: *Proceedings of the FEDCISIS 2012 Conference*, pp. 88–92 (2012)
14. Masud, M., Gao, J., Khan, L., Han, J., Thuraisingham, B.M.: Classification and novel class detection in concept-drifting data streams under time constraints. *IEEE Transactions on Knowledge and Data Engineering* 23(6), 859–874 (2011)
15. Ross, D.A., Lim, J., Lin, R., Yang, M.: Incremental learning for robust visual tracking. *International Journal of Computer Vision* 77(1-3), 125–141 (2008)
16. Schölkopf, B., Smola, A.J.: Learning with kernels: support vector machines, regularization, optimization, and beyond. In: *Adaptive Computation and Machine Learning*. MIT Press (2002)
17. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2008)
18. Zmyslony, M., Krawczyk, B., Woźniak, M.: Combined classifiers with neural fuser for spam detection. In: Herrero, A., Snasel, V., Abraham, A., Zelinka, I., Baruque, B., Quintin, H., Calvo, J.L., Sedano, J., Corchado, E. (eds.) *International Joint Conference CISIS12-ICEUTE12-SOCO12 Special Sessions*. AISC, vol. 189, pp. 245–252. Springer, Heidelberg (2012)

Comparable Study of Statistical Tests for Virtual Concept Drift Detection

Piotr Sobolewski and Michał Woźniak

Abstract. In this paper we examine the possibilities of using popular univariate statistical tests for discovering virtual concept drift in the stream of multidimensional data. Three popular methods are evaluated with different generalization approaches both on simulated and real data and compared by the specificity and sensitivity scores.

1 Introduction

In the machine learning literature a recently popular area of research due to the evolution of internet and expansion of decision making technology are the systems designed for classifying data streams [2]. Data streams are sources which continuously generate data, e.g. shopping trends, stock market, weather control, surveillance systems or health care. Classification in these areas is often hindered by various factors which cause undesirable changes in the data classification rules. Such phenomenon is called concept drift [12] and it is usually categorized in the literature by the type of changes in data classification rules and their impetuosity.

Type of change categorizes concept drifts as virtual or real. Virtual concept drifts occur when the changes affect solely the prior data distribution and the real concept drift occurs when the change affects the class-conditional likelihoods and the prior distribution of input data patterns remains unchanged [18] [17].

Another way of categorizing concept drift is by the impetuosity of changes, namely as incremental (small continuous changes, following a certain trend) or sudden (abrupt occasional changes, often referred to as "concept shift") [21]. Literature often mentions also a third type, called gradual concept drift which is similar to the sudden concept drift in terms of the scale of a change, however with a slower pace,

Piotr Sobolewski · Michał Woźniak

Wrocław University of Technology ul. Wybrzeże Wyspiańskiego 27 50-370 Wrocław

e-mail: {piotr.sobolewski, michał.wozniak}@pwr.wroc.pl

having two active concept sources at once during the switch. Our work does not consider this type of concept drift.

In this paper we focus on the sudden virtual concept drift. Examples of the sudden concept drifts (real and virtual, for comparison) are presented in Fig. 1.

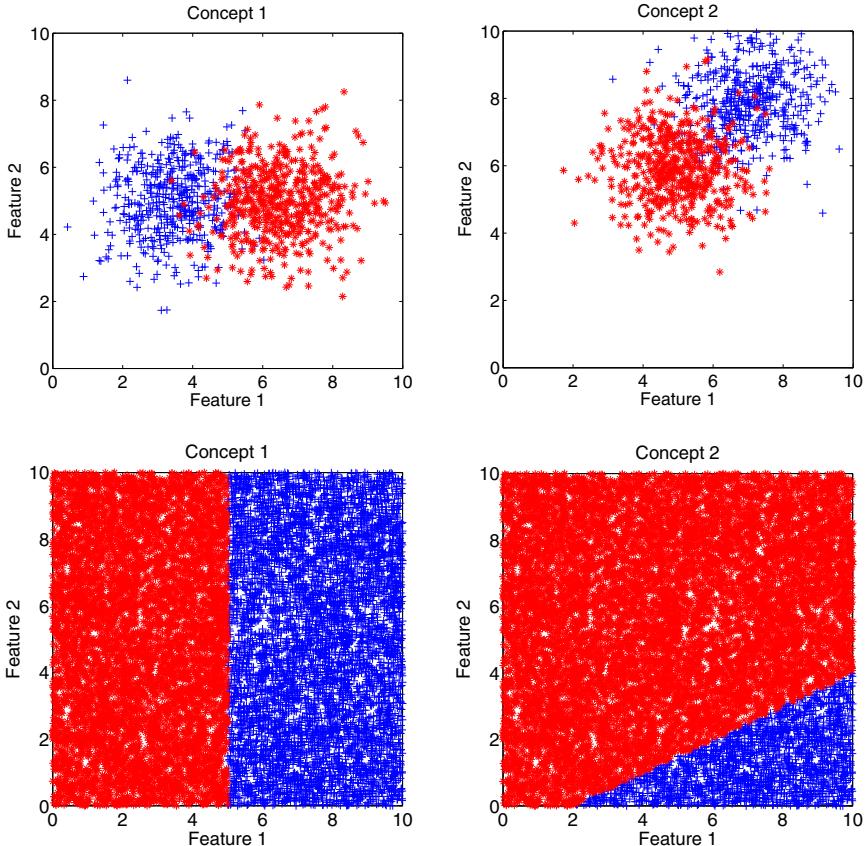


Fig. 1 Sudden virtual and sudden real concept drift scenarios

A change in the data classification rules is a threat for the decision making systems and therefore requires specific handling. Machine learning literature describes two general approaches to cope with concept drift [7]:

- Adapting a learner at the regular intervals without considering whether the changes have really occurred or not,
- First detecting the concept changes and then adapting a learner to them.

The second model distinguishes the classification module from the concept drift detector, allowing independent design of the two mechanisms. In the case of virtual

concept drift the detection can be performed by applying multivariate statistical tests on the data features, avoiding the cost of labeling data samples [10] [9].

2 Statistical Tests

In this paper we employ popular statistical tests in order to evaluate their efficiency as virtual concept drift detectors.

- **Kolmogorov-Smirnov test**

The two-sample Kolmogorov-Smirnov test [13] is non-parametric, as it makes no assumption about the distribution of data and therefore can be deployed on any data.

For the two-sample test, a Kolmogorov-Smirnov statistic is computed as

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \quad (1)$$

where $F_{1,n}$ and $F_{2,m}$ are the empirical distribution functions of samples computed as:

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \leq t\}, \quad (2)$$

where (x_1, \dots, x_n) are independent and identically distributed (i.i.d.) random variables laying in the real numbers domain with a common cumulative distribution function. The statistic is used to perform a KS-test to reject the null hypothesis at level α by computing:

$$\sqrt{\frac{nm}{n_m}} D_{n,m} > K_\alpha, \quad (3)$$

where K_α calculated from:

$$Pr(K \leq K_\alpha) = 1 - \alpha, \quad (4)$$

and K is a Kolmogorov distribution computed as:

$$K = \sup_{t \in [0,1]} |B(t)|, \quad (5)$$

$B(t)$ being the Brownian bridge [11].

In short, the Kolmogorov-Smirnov test compares the distributions of two samples by measuring a distance between the empirical distribution functions, taking into account both their location and shape.

- **Two-sample t-test** Two-sample t-test is one of the most popular tests used in economics and quality measures. It calculates the t-statistic on the basis of the means \bar{x}_1 , \bar{x}_2 , standard deviations s_1 , s_2 and the numbers of observations n_1 , n_2 in each sample by:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad (6)$$

which is then compared to the critical t -value taken from the significance tables, with regard to the number of the degrees of freedom, k :

$$k = \begin{cases} n_1 - 1, & \text{if } n_1 < n_2 \\ n_2 - 1, & \text{if } n_1 > n_2 \\ n_1 + n_2 - 2, & \text{if } n_1 = n_2 \end{cases} \quad (7)$$

The true outcome of the test means the rejection of the null hypothesis that both samples originate from the same distribution.

- **Wilcoxon Rank Sum test** Wilcoxon rank sum (also called Mann-Whitney-Wilcoxon) test [19] is a non-parametric alternative to the two-sample t-test, based solely on the order in which the observations from the two samples fall.

The test assumes, that all observations are independent from each other and can be ordered by their value, therefore if the test is performed on the data which are categorical, it has to be mapped to the numerical values.

The test ranks all observations regardless of which sample they are in by ordering them from the greatest to the lowest value. Then, a statistic is computed for each of the samples as:

$$U = R - \frac{n(n+1)}{2}, \quad (8)$$

where n is the sample size and R is the sum of the ranks in this sample. In order to reject the null hypothesis that both samples come from the same population, a lower value of U from both samples is chosen and consulted with the significance tables.

- **Wilcoxon Signed-rank test**

Another alternative to the T-test is a Wilcoxon Signed-rank test [19]. The test assumes, that the observations from both samples are independent and randomly paired, also observations need to be measured on an interval scale as the difference between the two has to be determined.

The test statistic is calculated on the basis of the ranks of the pairs, which are ordered by their absolute difference from the lowest to highest (pairs with difference equal to 0 are discarded) as:

$$W = \left| \sum_{i=1}^m [sgn(x_{2,i} - x_{1,i})R_i] \right| \quad (9)$$

The null hypothesis that the samples come from the same distribution is rejected if the z -score, calculated as:

$$z = \frac{W - 0.5}{\delta_W}, \quad (10)$$

$$\delta_W = \sqrt{\frac{N(N+1)(2N+1)}{6}}, \quad (11)$$

is higher than the critical z -value, taken from the significance tables. For small sample sizes, the W statistic value can be compared with a critical W value in the reference table [8].

- **Wald-Wolfowitz test**

The Wald-Wolfowitz test [16] [20] (also called the runs test) is based on the number of runs in a sequence of opposite elements, which is the number of groups of equal elements adjacent to each other. The test assumes, that the elements are independent and identically distributed.

The number of runs in a sequence of length N is a random variable, whose conditional distribution is approximately normal if the elements of the sequence are mutually independent. Given N observations, from which N_+ are positive and N_- are negative, the mean μ and variance σ^2 of the distribution can be calculated as follows:

$$\mu = \frac{2N_+N_-}{N} + 1, \quad (12)$$

$$\sigma^2 = \frac{(\mu - 1)(\mu - 2)}{N - 1}. \quad (13)$$

If the number of runs is significantly different than expected, the hypothesis of statistical independence can be rejected.

This test can be used to check whether two samples of observations follow the same distribution by comparing the values in one sample to the mean or median value of the other and marking the greater values with "+" and the lower with "-".

As every mentioned statistic is designed to operate on one-dimensional data, they need to be generalized to multivariate versions in order to work with the multidimensional data. There are many approaches described in the literature to achieve this [5] [3] [14] [6], however they are often complicated and therefore hard to implement.

2.1 Detecting Concept Drift

In the data stream, data samples are grouped into chunks of size 20, called the data windows. Detection is performed on the data window level, rather than on single samples.

The method of detecting concept drift with statistical tests is based on performing the statistical test on each of the data features in the group of samples drawn from the data stream and the reference dataset to determine whether the data comes from the same population. Concept drift is assumed if at least for one feature the test output is negative.

As some features may cause false-positive concept drift detections, we first discard the features, which may reduce the specificity of the test statistic. It is achieved in a very simple manner, by performing the tests on each of the data features on randomly drawn groups of samples from the reference dataset against and the rest of the dataset and discarding the ones, which produce too many false-positive detections.

A parameter λ is set heuristically and determines the false-positive error threshold, after which a feature should be discarded. For each test statistic s , a different set of features may be discarded, therefore for every s , a set of features F_s is saved and used for evaluation.

3 Experiments

The aim of experiments is to compare the efficiency in detecting virtual concept drift of the proposed statistical tests and the reference CNF Density Estimation test [4], designed specifically for concept drift detection.

3.1 Data

Experiments are performed both on the real and artificial data. The real data is borrowed from the UCI repository [1] and is specifically prepared to create concept drift in the data. In order to prepare a concept drifting environment, we follow the approach pioneered in [15] and deployed in [4]. The method can be described in two steps, first the data in dataset is ordered by classes from the most populated to the least populated, next class labels are removed and the data originating from each class is treated as data originating from different concepts. The data from the most populated class forms a reference dataset R and the data from the second most populated class forms a concept dataset, C .

The artificial datasets are two-class ten-dimensional Gaussian data, where concept drift occurs in three, five or ten data features, depending on the scenario. Concept drift is applied by changing the means of the class clusters in the affected data features.

3.2 Scoring and Results

Statistical tests are performed on the data windows and the detection results are stored for comparison. The scores are statistically validated with a t-test in order to determine a statistical significance of the differences in the results.

The scores are divided into two categories, specificity and sensitivity. Specificity score is a measure which represents how well a detector is able to detect that the concept has not changed, namely a low specificity score means many false-positive

errors. Sensitivity score on the other hand shows how well a detector is able to detect that the changes in the concept have occurred and a low sensitivity suggests a high false-negative error-rate.

The results are averaged scores of 100 test runs. Each test run is a different window of data generated for every dataset. The results of experiments are presented in Table 1.

Table 1 Results

Scenario	Sensitivity						Specificity					
	KS	Ttest	Wrank	Wsign	WW	CNF	KS	Ttest	Wrank	Wsign	WW	CNF
balance-scale	1.00	1.00	1.00	1.00	0.49	0.98	1.00	1.00	1.00	1.00	0.96	0.97
breast-w	1.00	1.00	1.00	1.00	1.00	0.78	1.00	1.00	1.00	1.00	1.00	1.00
car	1.00	1.00	1.00	0.92	0.28	0.74	1.00	1.00	1.00	1.00	0.96	0.96
credit-a	1.00	1.00	1.00	1.00	0.92	0.26	1.00	1.00	1.00	1.00	1.00	0.94
credit-g	0.93	1.00	1.00	0.92	0.60	0.09	1.00	1.00	1.00	1.00	0.99	0.97
haberman	0.86	0.94	0.96	0.68	0.14	0.10	1.00	1.00	1.00	0.98	0.98	0.98
heart-c	0.99	1.00	1.00	0.96	0.53	0.43	1.00	1.00	1.00	1.00	1.00	0.98
heart-statlog	1.00	1.00	1.00	1.00	0.88	0.54	1.00	1.00	1.00	1.00	1.00	1.00
ionosphere	1.00	1.00	1.00	1.00	0.88	0.00	1.00	1.00	1.00	1.00	1.00	1.00
kr-vs-kp	1.00	1.00	1.00	1.00	0.74	0.00	1.00	1.00	1.00	1.00	0.98	1.00
letter	1.00	1.00	1.00	1.00	1.00	0.42	1.00	1.00	1.00	1.00	1.00	0.97
mfeat-morph	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
nursery	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
optdigits	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
page-blocks	1.00	1.00	1.00	1.00	1.00	0.61	1.00	1.00	1.00	1.00	1.00	0.96
pendigits	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99
pima-diabetes	1.00	1.00	1.00	1.00	0.49	0.29	1.00	1.00	1.00	1.00	0.98	0.97
tic-tac-toe	0.59	0.61	0.77	0.45	0.36	0.46	0.99	1.00	0.99	1.00	0.99	0.98
vehicle	1.00	1.00	1.00	1.00	0.79	1.00	1.00	1.00	1.00	1.00	0.97	0.99
vote	1.00	1.00	1.00	1.00	1.00	0.24	1.00	1.00	1.00	1.00	1.00	0.97
waveform	1.00	1.00	1.00	1.00	0.99	0.57	1.00	1.00	1.00	1.00	1.00	0.96
yeast	0.80	0.87	0.87	0.74	0.20	0.00	1.00	1.00	1.00	1.00	0.97	1.00
art1	1.00	1.00	1.00	1.00	0.90	0.12	1.00	1.00	1.00	1.00	0.99	0.96
art2	1.00	1.00	1.00	1.00	0.94	0.12	1.00	1.00	1.00	1.00	1.00	0.96
art3	1.00	1.00	1.00	1.00	1.00	0.54	1.00	1.00	1.00	1.00	1.00	0.96

4 Conclusions and Future Work

First of all a high specificity score may suggest, that the statistical is efficient when concept drift does not occur, as it does not perform many false-negative detections. On the other hand, it could also mean that the test is not sensitive enough, what can be deduced by analyzing the sensitivity score. The best test would achieve high scores in both categories.

From the 6 evaluated test statistics, all achieve high specificity scores, therefore this score does not carry any valuable information in terms of efficiency comparison. It does prove however, that all tests are specific and could be safely used as detectors, which are required to make as little false-negative errors as possible.

The sensitivity scores show slightly more diversity, however they are also very similar. In most of the experiments, a statistical validation of the results did not show any difference among the first three statistics, namely the Kolmogorov-Smirnov, the t-test and the Wilcoxon rank sum tests. These three statistics achieved the highest score, surprisingly even higher than the CNF test, which was designed for concept drift detection purpose. An interesting fact is that the Wilcoxon Signed-rank test also achieved decent scores comparing with the CNF test, as this statistic requires the samples to have equal size, what limits the reference data to 20 randomly drawn samples. From 5 statistics only the Wald-Wolfowitz test failed as a concept detector.

Statistical tests have managed to achieve almost perfect scores both for real and artificial datasets, what suggests further research in more complicated and requiring environments. The scores are very motivating and suggest that using statistical tests as virtual concept drift detectors may be very efficient.

In the future, we are planning to expand the experiments and evaluate the statistical detectors on more challenging problems. Also, using different statistics for various scenarios might produce even better results, what could be achieved by deploying statistical detector ensembles which we plan to test and compare with more sophisticated concept drift detectors available in the literature.

Acknowledgements. This work is supported by the Polish National Science Center under a grant N N519 650440 for the period 2011-2014.

References

1. Newman, D.J., Asuncion, A.: UCI machine learning repository (2007)
2. Babcock, B., Babu, S., Datar, M., Motwani, R., Widom, J.: Models and issues in data stream systems. In: Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2002, pp. 1–16. ACM, New York (2002)
3. Corder, G.W., Foreman, D.I.: Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach, 1st edn. Wiley (May 2009)
4. Dries, A., Rückert, U.: Adaptive concept drift detection. Stat. Anal. Data Min. 2(5-6), 311–327 (2009)
5. Fasano, G., Franceschini, A.: A multidimensional version of the Kolmogorov-Smirnov test. Monthly Notices of the Royal Astronomical Society 225, 155–170 (1987)
6. Friedman, J., Rafsky, L.: Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests. In: The Annals of Statistics, pp. 697–717 (1979)
7. Greiner, R., Grove, A.J., Roth, D.: Learning cost-sensitive active classifiers. Artif. Intell. 139(2), 137–174 (2002)
8. Lowry, R.: Concepts and Applications of Inferential Statistics. Online (June 2007)

9. Markou, M., Singh, S.: Novelty detection: a review—part 1: statistical approaches. *Signal Processing* 83(12), 2481–2497 (2003)
10. Nishida, K., Yamauchi, K.: Detecting concept drift using statistical testing. In: Corrucole, V., Takeda, M., Suzuki, E. (eds.) DS 2007. LNCS (LNAI), vol. 4755, pp. 264–269. Springer, Heidelberg (2007)
11. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion (Grundlehren der mathematischen Wissenschaften), 3rd edn. Springer (December 2004)
12. Schlimmer, J.C., Granger Jr., R.H.: Incremental learning from noisy data. *Mach. Learn.* 1(3), 317–354 (1986)
13. Smirnov, N.V.: Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Stat.* 19, 279–281 (1948)
14. Sugiura, N.: Multisample and multivariate nonparametric tests based on U statistics and their asymptotic efficiencies. *Osaka J. Math.* 2, 385–426 (1965)
15. Vreeken, J., van Leeuwen, M., Siebes, A.: Characterising the difference. In: Berkhin, P., Caruana, R., Wu, X. (eds.) KDD, pp. 765–774. ACM (2007)
16. Wald, A.: Sequential Tests of Statistical Hypotheses. *The Annals of Mathematical Statistics* 16(2), 117–186 (1945)
17. Wang, S., Schlobach, S., Klein, M.C.A.: Concept drift and how to identify it. *J. Web Sem.* 9(3), 247–265 (2011)
18. Widmer, G., Kubat, M.: Effective learning in dynamic environments by explicit context tracking. In: Brazdil, P.B. (ed.) ECML 1993. LNCS, vol. 667, pp. 227–243. Springer, Heidelberg (1993)
19. Wilcoxon, F.: Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1(6), 80–83 (1945)
20. Wolfowitz, J.: On Wald's Proof of the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics* 20, 601–602 (1949)
21. Žliobaitė, I.: Learning under Concept Drift: an Overview. Technical report, Vilnius University, Faculty of Mathematics and Informatics (2009)

Part IV

Image Processing and Computer Vision

An Experimental Comparison of Fourier-Based Shape Descriptors in the General Shape Analysis Problem

Katarzyna Gościewska and Dariusz Frejlichowski

Abstract. The General Shape Analysis (GSA) is a problem similar to the typical recognition or retrieval of shapes, but it does not aim at the exact shape identification. The main goal is to find one or few most similar general templates, such as rectangle or triangle, for an investigated object. That allows for obtaining the most basic information about a shape. In this paper the experimental results on the application of three Fourier-based shape descriptors to the GSA problem are provided. The Euclidean distance is applied for measuring the dissimilarity between represented objects. In order to estimate the effectiveness of investigated shape descriptors the results of the experiments were compared with human benchmark results, collected by means of appropriate inquiry forms.

1 Introduction

The General Shape Analysis is a problem similar to the traditional recognition or retrieval of shapes, however there are some significant differences. In the GSA a small group of templates that are simple shapes (e.g. triangle or rectangle) and a larger group of more complicated test objects are utilized. By finding one or few most similar templates for each test object it is possible to determine general information about it, e.g. how rectangular or triangular it is. In order to estimate the similarity between particular objects, their representations obtained using shape description algorithms are matched by means of the template matching approach. The idea is to represent all shapes in the same way and to compare each test object with every template ones using a particular matching method (similarity or dissimilarity measure). The GSA problem is roughly depicted on Fig. 1.

Katarzyna Gościewska · Dariusz Frejlichowski

West Pomeranian University of Technology, Szczecin, Faculty of Computer Science and Information Technology, Żołnierska 52, 71-210, Szczecin, Poland

e-mail: {kgosciewska, dfrejlichowski}@wi.zut.edu.pl



Fig. 1 The illustration of the General Shape Analysis problem — which simple shape(s) is (are) the most similar to the test one?

The General Shape Analysis problem has been discussed in the literature and investigated with the use of different methods so far (e.g. [5]). It was for example successfully applied in the identification of stamp types [3]. The problem is very important nowadays since traditional paper documents are more often scanned to be stored on a hard drive, print or sent through the net. The main goal of this approach was to detect, localize and segment stamps from the electronic versions of paper documents in order to identify the stamp shape. The algorithm consisted of the following steps: colour conversion and separation, vertical and horizontal projection of pixel intensities and shape analysis using GSA approach. Since usually there are no specified stamps standards and stamps shapes are often irregular, the utilization of shape description algorithms in the general shape analysis approach outperformed traditional classification approach. In [8] a possible system-level hardware implementation of seven shape descriptors has been analysed.

Another two applications of GSA were described in [5]. General Shape Analysis can be applied in the multimedia databases where voice commands are used for shape retrieval (e.g. "find red rectangle") and can speed up the process of searching large databases by iteratively narrowing the group of the most possible matches — this can be considered as a coarse classification. In addition, the GSA problem can be applied for the analysis of biological objects development such as plants [9], and for generalized shape matching [2]. Triangular and rectangular aspects of shape have been used e.g. for the discrimination of fish [15], and the ellipse has been applied e.g. for modelling corneal shape [13] or microfossils [1].

Referring to the aforementioned applications, primarily to stamps identification, the General Shape Analysis approach is very useful and helpful especially if the data are incomplete or there is no information about the amount of missing data. This problem has been relatively frequently considered in the literature — the authors of [10] investigated the problem of incomplete image registration while in [11] the authors applied the evolution-based shape matching. There are also some publications describing another approach to the exploration of the general shape, e.g. works of Paul Rosin, who investigated the ability of various shape measures to define the degree to which a processed shape differs from a general template. The author proposed his own algorithms and explored some traditional methods for measuring single shape properties. In [16] three new approaches for measuring rectangularity were proposed and tested together with the standard minimum bounding rectangle method. Several ellipticity, triangularity and rectangularity shape descriptors have been described and evaluated in [17] using both real and synthetic data. Wide

variety of global shape measures has been reviewed in [18]. Additionally, new method for measuring squareness was presented in [19].

In this paper, for the first time, the GSA problem has been investigated with the use of three various Fourier-based shape descriptors. The main goal was to study their performance and to determine the most important characteristics of each description algorithm that have significant influence on final accuracy of the results. The reason for utilizing Fourier Descriptors (FD) is associated with valuable properties of Fourier transform, especially the generalization of the represented object. FD captures both global and local features in spectral domain and by reducing the size of absolute spectrum it is possible to obtain concise and compact shape representation.

The rest of the paper is organized as follows. The second section presents three applied shape descriptors that are two-dimensional Fourier Descriptor (2D FD), Generic Fourier Descriptor (GFD) and UNL-Fourier (UNL-F) descriptor. The third section contains the experimental conditions and results. The last one concludes the paper.

2 Fourier-Based Shape Descriptors

Fourier-based shape descriptors are well-known and widely utilized for various applications thanks to the useful properties of Fourier transform. FDs are easy to compute and compact what simplifies the matching process. Moreover, the obtained representations are robust to noise and invariant to scale and translation. In the paper only two-dimensional Fourier transform is utilized and various sizes of absolute spectrum subpart are investigated if possible. A brief description of selected shape descriptors is provided in the following subsections.

2.1 Two-Dimensional Fourier Descriptor

Two-dimensional Fourier Descriptor is applied to a region shape representation (with its interior). It has a form of a matrix with absolute complex values that can be derived using the following formula [12]:

$$C(k, l) = \frac{1}{HW} \left| \sum_{h=1}^H \sum_{w=1}^W P(h, w) \cdot e^{(-i\frac{2\pi}{H}(k-1)(h-1))} \cdot e^{(-i\frac{2\pi}{W}(l-1)(w-1))} \right|, \quad (1)$$

where:

H, W — height and width of the image in pixels,

k — sampling rate in vertical direction ($k \geq 1$ and $k \leq H$),

l — sampling rate in horizontal direction ($l \geq 1$ and $l \leq W$),
 $C(k, l)$ — value of the coefficient of discrete Fourier transform in the coefficient matrix in k row and l column,
 $P(h, w)$ — value in the image plane with coordinates h, w .

The above-mentioned Fourier transform will be applied as a final step in the two methods described below.

2.2 UNL-F Fourier

The UNL-Fourier descriptor is composed of UNL (named after Universidade Nova de Lisboa) descriptor and Fourier transform as an additional step. The first transform was successfully applied for example in recognition of airplane silhouettes [4], erythrocyte types for the needs of automatic diagnosis [6] and signs extracted from car license plates [7]. The UNL utilizes complex representation of Cartesian coordinates for points and parametric curves in discrete manner [14]:

$$z(t) = (x_1 + t(x_2 - x_1)) + j(y_1 + t(y_2 - y_1)), \quad t \in (0, 1), \quad (2)$$

where $z_1 = x_1 + jy_1$ and $z_2 = x_2 + jy_2$ are complex numbers. Then, the centroid O is calculated [14]:

$$O = (O_x, O_y) = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n y_i \right), \quad (3)$$

and the maximal Euclidean distance between contour points and centroid is found [14]:

$$M = \max_i \{ \|z_i(t) - G\| \} \quad \forall i = 1 \dots n \quad t \in (0, 1). \quad (4)$$

The discrete version of new coordinates can be derived using the following formula [14]:

$$\begin{aligned} U(z(t)) = & \frac{\|(x_1 + t(x_2 + x_1) - G_x) + j(y_1 + t(y_2 + y_1) - G_y)\|}{M} + \\ & + j \times \arctan \left(\frac{y_1 + t(y_2 - y_1) - G_y}{x_1 + t(x_2 - x_1) - G_x} \right). \end{aligned}$$

The parameter t is discretized in the interval $[0, 1]$. New coordinate values are put into a matrix in which rows represent distances from the centroid and columns — the angles. This results in an image containing unfolded shape contour in polar coordinates. Thanks to this, two-dimensional Fourier transform can be applied.

2.3 Generic Fourier Descriptor

Generic Fourier Descriptor is a region-based FD that utilizes the transformation to the polar coordinate system. All pixel coordinates from original region shape image are transformed into polar coordinates and new values are put to a rectangular Cartesian image [20]. The row elements correspond to distances from centroid and the columns to 360 angles. The result has a form of an image and as in the previous case two-dimensional Fourier transform is applied.

3 Experimental Conditions and Results

Three experiments on the General Shape Analysis were carried out — 2D FD, GFD and UNL-F descriptor were explored. The Euclidean distance was used in order to measure the dissimilarity between test and template objects. The experiments were performed using 50 various shapes that were 200×200 pixel size binary images. The shapes were divided into two groups — 10 of them were general templates and the rest were test objects (see Fig. 2). General templates are the simplest geometric figures that represent general shape features, e.g. the rectangle is useful in evaluating the rectangularity, circle for circularity, and so on. All experiments were performed in the same way: every shape was represented by means of particular shape descriptor. Then, on the basis of the template matching approach, Euclidean distance between each test object and every template was estimated. Three templates with the smallest dissimilarity values indicated the most similar shapes, i.e. the shapes which vary least from the test object, thus the major features of the object are obtained. Various sizes of absolute spectrum subpart for 2D FD and GFD were explored as well.

In order to recognize the overall effectiveness of the descriptors, accuracy measure was used. Accuracy was calculated for particular indications made for each possible representation size as a percentage proportion of correctly selected templates to the number of all selections in the experiment (equal to 40 test objects). A template is correctly selected when there is a coincidence between experimental results and results provided by humans through inquiry forms that concerned the same GSA task. Shapes that were indicated most frequently in the inquiries were the most similar templates and were treated as a benchmark result. However, for some shapes the number of indications for the first and the second template was very similar. Therefore, both shapes were considered as the most similar ones and compared separately with the first shape indicated in the experiments. In conclusion, for each shape representation of a specified size, the accuracy values were estimated for the 1st, 2nd and 3rd indication out of three most similar templates and for one most similar template separately.

The first experiment used two-dimensional Fourier Descriptor. It was performed five times separately for various parts of absolute spectrum that were $n \times n$ blocks taken from the top left corner of the coefficient matrix, and n was equal to 2, 5, 10,

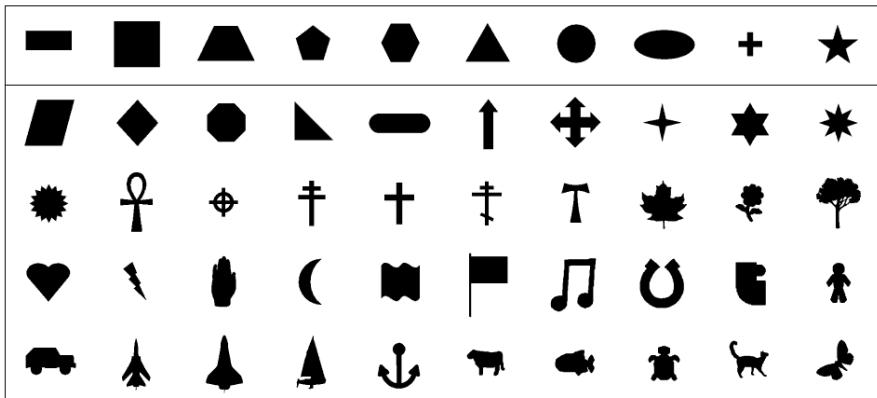


Fig. 2 Shapes used in the experiments – 10 templates in the first row and 40 test objects [5]

Table 1 Accuracy values for the first experiment utilizing various sizes of FD representations

Subpart size	Three most similar templates			One most similar template
	1st indication	2nd indication	3rd indication	
2×2	30,0%	20,0%	25,0%	32,5%
5×5	25,0%	17,5%	35,0%	27,5%
10×10	25,0%	17,5%	20,0%	30,0%
25×25	25,0%	17,5%	22,5%	30,0%
50×50	25,0%	17,5%	20,0%	30,0%

25 and 50. Accuracy values achieved for one and three most similar templates are provided in Table 1.

The best results were obtained for the smallest subpart size and were equal to 30%, 20% and 25% for the 1st, 2nd and 3rd indication out of three most similar templates, and 32,5% for one most similar template. It means that 2×2 subpart contains the most important and general information about a shape and is sufficient to discriminate one shape from another. The most efficient results are pictorially presented in Fig. 3.

The second experiment investigated UNL-F descriptor and the original size of the representation was reduced to 10×10 subpart. UNL-F utilizes only contour information but transformed into polar coordinates and represented on the image plane. The results for the 1st, 2nd and 3rd indication out of three most similar templates were equal to 15%, 10% and 7,5% respectively. For one most similar template the accuracy was equal to 22,5%, what gives 7,5% improvement. This means that in the experiments the appropriate shapes were indicated but in some cases in the different order than humans did. Pictorial results are provided in Fig. 4.

The third experiment utilized GFD and various sizes of absolute spectrum subpart. Accuracy values obtained for the 5×5 and larger blocks as well as for the

Test object	Templates		
	1 st	2 nd	3 rd
1	■	▲ ■ □ ○	
2	◆	○ ▲ △ ♦	
3	●	● ○ △ ▲	
4	△	▲ △ ○ ▶	
5	—	— ○ ▲	
6	↑	+ ★ △	
7	+	▲ ○ —	
8	Ω	★ + ▲	
9	◆	+ ★ △	
10	†	+ ★ △	
11	†	+ ★ △	
12	†	+ ★ △	
13	T	+ ★ △	
14	+	+ ★ △	
15	★	○ ▲ —	
16	*	★ △ ▲	
17	●	▲ ○ —	
18	▼	★ ○ ▲	
19	◀	★ ○ ▲	
20	★	★ ○ ▲	
21	λ	★ + ▲	
22	●	★ + △	
23	●	★ △ ▲	
24	●	● ○ ▲	
25	●	★ ○ ▲	
26	●	▲ ○ —	
27	●	○ △ ○ ▲	
28	↖	+ ★ △	
29	○	▲ ○ △	
30	○	★ + △	
31	■	● ▲ ○ △	
32	■	● ○ △ ▲	
33	■	★ ▲ ○ △	
34	○	▲ — ○ △	
35	○	○ △ ● ▲	
36	○	— ▲ ○ △	
37	▲	★ ○ △	
38	▲	▲ ○ △	
39	▲	○ ▲ ○ △	
40	○	★ + ▲	

Fig. 3 Results of the experiment utilizing 2×2 block of two-dimensional FD representation

Test object	Templates		
	1 st	2 nd	3 rd
1	■	○ — □ ▲	
2	◆	□ ○ — ○	
3	●	● ○ △ ▲	
4	△	▲ ○ △ ▲	
5	—	— ○ ▲	
6	↑	— ○ ▲	
7	+	+ ■ ★	
8	Ω	▲ ★ —	
9	◆	□ — ○	
10	†	▲ ▲ —	
11	†	▲ — ▲	
12	†	▲ ▲ —	
13	T	▲ ▲ ○	
14	+	+ ■ —	
15	★	▲ —	
16	*	+ ★ ■	
17	●	★ ○ ■	
18	▼	▲ ○ —	
19	◀	▲ ○ —	
20	★	▲ ○ —	
21	λ	▲ ★ —	
22	●	▲ — ★	
23	●	— ▲ ▲	
24	●	★ ▲ ○	
25	●	▲ ★ ○	
26	●	▲ ○ △	
27	●	▲ ○ ○ △	
28	↖	— ▲ ○ △	
29	○	○ ▲ ○ —	
30	○	▲ ○ —	
31	■	■ ○ ○ —	
32	■	▲ ▲ ○ —	
33	■	▲ — ▲ ○	
34	○	○ ▲ ○ △	
35	○	▲ ○ ○ —	
36	○	▲ ▲ ○ —	
37	▲	▲ ▲ ○ ○	
38	▲	▲ ▲ ○ ○	
39	▲	○ ▲ ○ ○	
40	○	+ — ▲	

Fig. 4 Results of the experiment utilizing UNL-F descriptor

whole representation were the same. Therefore, only 2×2 subpart and the original representation were taken into consideration. The accuracy values are provided in Table 2.

Better results were achieved using the smallest subpart and were equal to 32,5%, 22,5% and 2,5% for the 1st, 2nd and 3rd indication out of three most similar

Table 2 Accuracy values of the third experiment for various sizes of GFD representations

Subpart size	Three most similar templates			One most similar template
	1st indication	2nd indication	3rd indication	
2×2	32,5%	22,5%	2,5%	35,0%
200×200	27,5%	20,0%	5,0%	30,0%

Test object	Templates			Test object	Templates			Test object	Templates			Test object	Templates			
	1 st	2 nd	3 rd		1 st	2 nd	3 rd		1 st	2 nd	3 rd		1 st	2 nd	3 rd	
1	■	▲	□	○	+	★	—	21	●	★	+	—	31	■	●	▲
2	◆	◆	◆	●	+	★	—	22	●	★	+	▲	32	—	—	★
3	●	●	●	▲	+	★	—	23	●	★	+	—	33	—	+	★
4	◀	—	★	▲	+	★	—	24	●	●	●	●	34	●	+	★
5	—	▲	◆	—	★	▲	●	25	●	★	—	+	35	●	◆	●
6	↑	+	★	—	*	★	▲	26	●	—	★	▲	36	●	—	▲
7	+	▲	◆	—	*	●	◆	27	●	▲	—	●	37	●	★	—
8	♀	+	★	—	*	●	—	28	●	+	★	—	38	●	—	▲
9	◆	+	★	—	*	●	◆	29	●	▲	◆	●	39	●	—	★
10	†	+	★	—	*	●	▲	30	●	+	★	—	40	●	+	★

Fig. 5 Results of the experiment utilizing 2×2 block of GFD representation

templates respectively and 35,0% for one most similar template. Pictorial results of the experiment with the highest accuracy values are provided in Fig. 5.

The accuracy values of the last experiment are similar to those obtained in the experiment utilizing 2D FD, especially for the 1st and the 2nd indication. Despite the fact that GFD and UNL-F methods are similar and require a transformation to polar coordinate system, the experiment with the GFD gave much better results. It means that utilizing region shapes instead of contour ones brings more information about a shape and results in higher effectiveness. Moreover, the size of extracted absolute spectrum subpart for region shapes has an influence on results as well. The smaller the subpart size the better the result.

4 Conclusions

In the paper the problem of the General Shape Analysis was described. It is similar to the traditional recognition or retrieval of shapes but the identification is not performed. For every processed shape one or few most similar general templates are

found. It is usually done on the basis of the template matching approach. Templates are simple shapes and by indicating which is the most similar the basic information about a shape is obtained, e.g. how rectangular or triangular it is.

During the experiments, three Fourier-based shape descriptors were explored, namely two-dimensional Fourier Descriptor, Generic Fourier Descriptor and UNL-Fourier. Each experiment was performed on the basis of template matching approach and with the use of Euclidean distance as dissimilarity measure. The accuracy of the experimental results was estimated for each possible representation size as a percentage proportion of correctly selected templates to the number of all selections made separately for the 1st, 2nd and 3rd indication. The template selected in the experiment was correct only if it was consistent with the shape chosen by humans in the inquiries. The best results were obtained in the experiment utilizing Generic Fourier Descriptor and 2×2 block of absolute spectrum. The accuracy was equal to 32,5%, 22,5% and 2,5% for the 1st, 2nd and 3rd indication out of three most similar templates respectively and 35% for one most similar template.

The comparison of investigated algorithms gave several conclusions. First, better results in the General Shape Analysis using the Fourier transform-based algorithms can be obtained for the region shape instead of the contour one. Moreover, it is more efficient when the shape with its interior is transformed into polar coordinate system. While processing the region shapes the size of absolute spectrum subpart has to be taken into consideration. It can significantly influence the final results.

Obviously, future work can be done in order to find a method that will result in higher accuracy values. The approaches provided by P. Rosin can be introduced to GSA problem in the future as well.

References

1. Brechner, S., Ade, F.: An Automatic System for the Classification of Microfossils. In: Proceedings of the 11th Scandinavian Conference on Image Analysis, vol. 2, pp. 825–832 (1999)
2. Dryden, I.L.: General shape and registration analysis. Stochastic Geometry: Likelihood and Computation (1997)
3. Forczmański, P., Frejlichowski, D.: Robust Stamps Detection and Classification by Means of General Shape Analysis. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part I. LNCS, vol. 6374, pp. 360–367. Springer, Heidelberg (2010)
4. Frejlichowski, D.: An Algorithm for Binary Contour Objects Representation and Recognition. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2008. LNCS, vol. 5112, pp. 537–546. Springer, Heidelberg (2008)
5. Frejlichowski, D.: An Experimental Comparison of Seven Shape Descriptors in the General Shape Analysis Problem. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010. LNCS, vol. 6111, pp. 294–305. Springer, Heidelberg (2010)

6. Frejlichowski, D.: Pre-processing, Extraction and Recognition of Binary Erythrocyte Shapes for Computer-Assisted Diagnosis Based on MGG Images. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part I. LNCS, vol. 6374, pp. 368–375. Springer, Heidelberg (2010)
7. Frejlichowski, D.: Analysis of four polar shape descriptors properties in an exemplary application. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part I. LNCS, vol. 6374, pp. 376–383. Springer, Heidelberg (2010)
8. Frejlichowski, D.: Analysis of possible system-level hardware implementation of selected shape description algorithms. *Journal of Theoretical and Applied Computer Science* 6(4), 51–58 (2012)
9. Gleissberg, S.: Comparative Analysis Of Leaf Shape Development In *Eschscholzia California* And Other Papaveraceae-Schscholzioideae. *Am J. Bot.* 91(3), 306–312 (2004)
10. Gut, P., Chmielewski, L., Kukołowicz, P., Dąbrowski, A.: Edge-based robust image registration for incomplete and partly erroneous data. In: Skarbek, W. (ed.) CAIP 2001. LNCS, vol. 2124, pp. 309–316. Springer, Heidelberg (2001)
11. Krupiński, R., Mazurek, P.: Electrooculography Signal Estimation by Using Evolution-Based Technique for Computer Animation Applications. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part I. LNCS, vol. 6374, pp. 139–146. Springer, Heidelberg (2010)
12. Kukharev, G.: Digital Image Processing and Analysis. SUT Press, Szczecin (1998) (in Polish)
13. Lindsay, R., Smith, G., Atchison, D.: Descriptors of corneal shape. *Optometry Vision Sci.* 75(2), 156–158 (1998)
14. Rauber, T.W.: Two Dimensional Shape Description. Technical report: GR UNINOVA-RT-10-94. Universidade Nova de Lisboa, Lisboa, Portugal (1994)
15. Richards, R.A., Esteves, C.: Use of scale morphology for discriminating wild stocks of Atlantic striped bass. *T. Am Fish Soc.* 126(6), 919–925 (1997)
16. Rosin, P.L.: Measuring Rectangularity. *Mach. Vision Appl.* 11(4), 191–196 (1999)
17. Rosin, P.L.: Measuring Shape: Ellipticity, Rectangularity, and Triangularity. *Mach. Vision Appl.* 14(3), 172–184 (2003)
18. Rosin, P.L.: Computing Global Shape Measures. In: Chen, C.H., Wang, P.S.P. (eds.) *Handbook of Pattern Recognition and Computer Vision*, 3rd edn., pp. 177–196 (2005)
19. Rosin, P.L., Žunić J: Measuring Squareness and Orientation of Shapes. *J. Math. Imaging Vis.* 39, 13–27 (2011)
20. Zhang, D., Lu, G.: Shape-Based Image Retrieval Using Generic Fourier Descriptor. *Signal Process-Image* 17(10), 825–848 (2002)

Extraction of the Foreground Regions by Means of the Adaptive Background Modelling Based on Various Colour Components for a Visual Surveillance System

Dariusz Frejlichowski, Katarzyna Gościewska, Paweł Forczmański,
Adam Nowosielski, and Radosław Hofman

Abstract. Intelligent monitoring systems based on visual content analysis are often composed of three main modules — background modelling, object extraction and object tracking. This paper describes a method for adaptive background modelling utilizing Gaussian Mixture Models (GMM) and various colour components. The description is based on the experimental results obtained during the development of the SmartMonitor — an innovative security system based on video content analysis. In this paper the main characteristics of the system are introduced. An explanation of GMM algorithm and a presentation of its main advantages and drawbacks is provided. Finally, some experimentally obtained images containing foreground regions extracted with the use of various background models are presented.

1 Introduction

Intelligent monitoring systems utilizing video content analysis (VCA) algorithms are usually expensive and narrowly targeted to the specific group of users. For example, an advanced surveillance system Bosch IVA 4.0 [15] was designed to support operators of CCTV monitoring and to be applied on larger areas or in public buildings, such as airports. The system described in [10] utilizes Hidden Markov Models to analyse human behaviour, but it requires the participation of a qualified employee and the preparation of a large database for the learning process [10]. In the Smart-Monitor system human involvement will be reduced to the minimum, especially due to feature-based methods and simple calibration. Moreover, in monitoring systems

Dariusz Frejlichowski · Katarzyna Gościewska · Paweł Forczmański · Adam Nowosielski
West Pomeranian University of Technology, Szczecin, Faculty of Computer Science and
Information Technology, Żołnierska 52, 71-210, Szczecin, Poland
e-mail: {dfrejlichowski, pforczmanski, anowosielski}@wi.zut.edu.pl

Katarzyna Gościewska · Radosław Hofman
Smart Monitor, sp. z o. o., Niemierzyńska 17a, 71-441, Szczecin, Poland
e-mail: {katarzyna.gosciewska, radekh}@smartmonitor.pl

with VCA functionality the process of localization and extraction of moving objects occurring in video streams is usually based on background subtraction methods that utilize various background models. In the literature one can find background models that are static [6], averaged in time [3] or built adaptively with the use of Gaussian Mixture Models (e.g. [9, 11, 13]). The last category is very valuable and useful during the analysis of real scenes that are very changeable in time. Hence it was applied in the SmartMonitor system.

SmartMonitor is an innovative visual surveillance system that is being currently developed to meet the needs of individual clients who want to ensure the safety of their assets and surrounding areas. The system will operate in four predefined scenarios and utilize video content analysis algorithms in order to detect and to react to every change in a pre-specified way. The utilization of VCA methods will allow for the elimination of a large part of human involvement, characteristic for traditional monitoring systems. Only the calibration process will require human interaction — it is one of the crucial system advantages. Every user will be able to set individual safety rules and to adjust system sensitivity degree to the actual requirements. Moreover, SmartMonitor will utilize commonly available and affordable hardware, i.e. personal computer and digital camera(s). The system is going to be an inexpensive solution as well.

In the SmartMonitor system we adopted general scheme of background subtraction process that is depicted in Fig. 1 and includes: image pre-processing (e.g. contrast improvement or image compression), background modelling, foreground localization and data validation. Foreground image is obtained by subtracting background image from the currently processed frame. Data validation stage allows for artefacts removal and verifies whether the foreground image region is an object of interest (OOI) for the system. Every OOI is supposed to be a coherent region larger than a specified size.

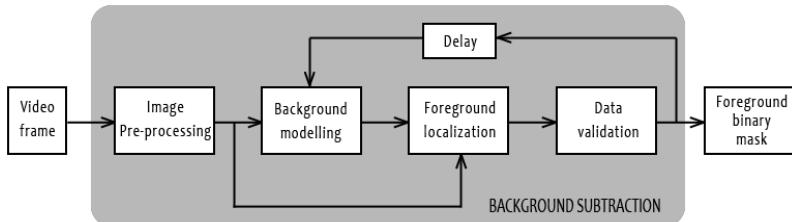


Fig. 1 Proposed flow diagram presenting the generic algorithm for background subtraction (based on [11])

The SmartMonitor system has been described and investigated in [2, 4, 5] so far. This paper is focused on two of the aforementioned steps, i.e. background modelling and foreground localization. Despite the fact that the system is currently at the developmental stage, some experiments have been already performed. Various colour components applied for building the background models were investigated and evaluated. The evaluation was carried out on the basis of the overall precision

and accuracy of localized foreground regions. The main goal for investigating different colour components is to find the most optimal working system parameters in each scenario.

The rest of the paper is organized as follows: the second section describes main characteristics of the system and presents system working scenarios. In the third section the explanation of Gaussian Mixture Model and its possible modifications is provided. The fourth section contains exemplary results on utilizing various background models in the process of background subtraction. The last section concludes the paper.

2 Main Properties of the Developed System

SmartMonitor will be able to operate in four predefined scenarios. Each scenario is a combination of intended or predicted situations, associated with different scenes and environments. This results in various system working conditions and different groups of performed actions. In this paper we focus on three out of four scenarios, i.e. scenario A — home/surrounding protection against unauthorized intrusion, scenario B — supervision of ill person and scenario C — crime detection. The fourth scenario, smoke and fire detection, is not taken into consideration — the extraction of foreground regions is not performed. There are some actions and conditions common for scenarios A, B and C, i.e. movement detection, object tracking and possibility of region limitation. The detection of any movement would activate the alarm. The behaviour of moving objects would be analysed using trajectories. Movement detection and object tracking can be performed on the limited region which could additionally increase system effectiveness.

The monitoring system working in scenario A is very similar to the traditional one — any detected movement induces the alarm. However, in the SmartMonitor, basic classification is performed in order to verify whether the moving object is a human or not. Then, only particular objects, e.g. of a specified size, will be taken into consideration. The system working in scenario A should be active at night as well. Moreover, weather conditions that can influence captured images have to be taken into account. Sudden changes in lighting could affect image colour and increase the difference between foreground and background images. In result the foreground region will include false detections. Sample frames presenting scenes appropriate for scenario A are depicted in Fig. 2.

Scenario B is suitable for the system working inside the building, mainly in houses, flats or private apartments as well as nursing homes e.g. for elderly or sick people. It focuses on incident detection, especially resulting in changes in object shape and trajectory. The main goal is to recognize faint and fall as sudden change in proportions and lack of movement lasting for pre-specified time. It is important to localize the exact object region without false detections as it is crucial for shape analysis and recognition. Fig. 3 contains sample frames presenting a person who enters the scene, falls after some time and lies still.



Fig. 2 Two sample frames presenting running objects in a garden — basic human/not-human classification exclude the alarm activation when dogs are detected (Scenario A)



Fig. 3 Sample frames presenting the simulation of ill person fainting (Scenario B)

Scenario C is very similar to the previous one, however the system reacts to different types of changes, such as raising hands up or unusual trajectory. System working in scenario C is intended to be implemented e.g. in offices, shops or other small enterprises. It would be helpful in the situation when a threatened person cannot activate the alarm — in that case the system will detect suspicious behaviour and be able to send information and images to appropriate services. Fig. 4 contains two sample frames presenting a person with hands raised up. Different lighting distributions are noticeable — this can cause problems while building a background model depending on various colour components.



Fig. 4 Sample frames presenting the simulation of a crime scene (Scenario C)

3 Adaptive Background Modelling Algorithm Based on Gaussian Mixture Models

Performing background subtraction, which identifies moving objects, requires the development of good background model. Background model has to be both robust to any environmental changes in the scene, such as swinging leaves, rain, illumination

changes or shadows, and sufficiently sensitive to identify all objects of interest. It is important to find a compromise between these two conditions, i.e. a model that will adapt to various and unexpected changes in real scenes. Therefore, the Gaussian Mixture Model [11, 12] was chosen to meet the mentioned requirements.

GMM is an adaptive background modelling technique that models each pixel as a mixture of Gaussians and utilizes an on-line approximation for model update. Background model quickly adapts to the changes in the scene and allows detecting any moving object. GMM have been tested and evaluated as a reliable, stable and real-time outdoor tracker that is able to deal with long-term changes and influence of light [12]. It tracks multiple Gaussian distributions at the same time and maintains a density function for every pixel. The model is parametric and can be adaptively updated with every consecutive frame without keeping a large video frames buffer [11]. GMM algorithm can use various colour information as an input. The authors of the method [12] tested it with the use of grayscale, RGB and HSV colour spaces (pixel values can be e.g. vectors for colour images or scalars for grayscale values). The description of the GMM method, based on [12] and [11], is provided below.

The sequence of video frames, denoted as $\{X_1, \dots, X_t\}$, is an input for the GMM algorithm. X is a single frame, t is a frame index, $M \times N \times K$ defines a frame size — M and N determine spatial resolution and K is the number of RGB colour space components. The probability P of a particular $x_t(m, n)$ pixel value occurrence is defined by G Gaussian distributions and given by the formula:

$$P(x_t(m, n)) = \sum_{i=1}^G \omega_{i,t} * \eta(x_t(m, n), \mu_{i,t}, \Sigma_{i,t}). \quad (1)$$

Gaussian distributions are determined by following parameters: $\mu_{i,t}$ is the mean value of the i^{th} Gaussian in the mixture at time t , $\Sigma_{i,t}$ is the covariance matrix of the i^{th} Gaussian in the mixture at time t , $\omega_{i,t}$ is an estimate of the weight, σ is the standard deviation and η is the Gaussian probability density function.

Background model can be initialized with random values or with the first frame obtained from the video stream. The utilization of random values requires more time for the system to adapt pixel values to the current conditions in the scene. However, the longer the model is being adapted, the more information is taken into consideration and the model adjustment is more accurate. The initialization with the first frame gives a ready model, but it can be affected by the objects that do not belong to the background region. Therefore, this solution is only suitable when there are no foreground objects in the first frame. In both cases, the initial model is then iteratively adapted with the use of GMM algorithm. In each iteration every pixel from the current frame is analysed and the background model is modified. The modification depends on the comparison between $x_t(m, n)$ pixel value from X_t frame and model value μ — the distance d is calculated as follows:

$$d = |(x_t(m, n) - \mu)|. \quad (2)$$

If the distance value $d > c * \sigma$, where c is a constant, then the particular pixel do not belong to the background image. In that case the least probable distribution is replaced by a distribution that includes current pixel value as its mean value. Otherwise, if $d < c * \sigma$, the model is adapted including new estimated value using following formulas:

$$\mu_t = (1 - \rho) \mu_{t-1} + \rho x_t(m, n), \quad (3)$$

$$\sigma_t^2 = (1 - \rho) \sigma_{t-1}^2 + \rho (x_t(m, n) - \mu_t)^T (x_t(m, n) - \mu_t), \quad (4)$$

where $\rho = \alpha \eta(x_t(m, n) | \mu_{t-1}, \sigma_{t-1})$ and α is a learning constant. Subsequently, estimates of the weight in particular distributions are modified:

$$\omega_t = \begin{cases} (1 - \alpha) \omega_{t-1} + \alpha & \text{if new value was matched,} \\ (1 - \alpha) \omega_{t-1} & \text{otherwise.} \end{cases}$$

In order to estimate the current background image, the modelled distributions are sorted by decreasing values of the ω/σ ratio and put on a list. The most probable distributions with high weight and low standard deviation are placed at the top of a list and the least probable are on the bottom — these ones can be substituted by new distributions. New estimated scene model is created by B first distributions that meet following condition:

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^b \omega_k > T \right), \quad (5)$$

where T is the proportion of the data that should be accounted by the background. When new object occurs in the scene and stays stationary long enough to become a part of the background, the existing model is not destroyed — only the least probable representation of it is removed. The original background remains with the same μ and σ^2 , but lower ω , and when the object moves again it is possible to quickly include appropriate distributions describing previous background model.

Despite the mentioned advantages there are also some drawbacks. The first problem is associated with the number of Gaussian distributions describing each pixel — it was stated in [14] that it should be equalled to 3 to 5 models or modified adaptively depending on the actual needs. The next problem is associated with the update equation. In [8] the GMM algorithm modification that concerns the reinvestigation of update equation was proposed. This can allow the system to learn faster, adapt more effectively to changing environment and detect shadow areas. Unfortunately, the GMM algorithm does not distinguish between moving objects and moving shadows. In order to solve the problem the colour space that can separate chromatic and intensity components has to be utilized. Authors of [1] investigated the usage of the H component from the HSV colour scheme, which corresponds more closely to human perception. It was experimentally found that shadow darkens the region while the hue does not vary too much [1]. Moreover, the GMM algorithm is characterized

by a low robustness to sudden changes in lighting what results in numerous false detections. Therefore, the utilization of colour features and gradients was proposed in [7].

4 Conditions of the Experiments and Exemplary Results

In this section the experimental results on utilizing Gaussian Mixture Models basing on various colour components are described. The main goal of the experiments was to find the best working conditions for the system. It would be achieved by the selection of the most efficient solution: it is expected that the model will allow for the accurate detection of moving objects and that the extracted regions will not be affected by false detections or it will be possible to distinguish artefacts from an actual OOI. The accuracy of detected regions was evaluated by a user — the visual evaluation of final foreground binary masks in comparison to the particular input images was performed. If the foreground mask contained only the region of OOI or additional smaller regions that would be eliminated on the basis of the minimum size condition, then the result was considered as a good result. The experiments were performed with the use of test database containing 32 video sequences that simulated situations appropriate for each examined scenario. The test videos were recorded on various locations, inside as well as outside the buildings, in different lighting conditions and in the presence of various moving objects. Total duration time of all sequences amounted to more than 45 minutes. Some of the experimental results for scenario A, B and C are provided in Fig. 5–7. Each figure contains following subimages: (a) input frame, (b) result for RGB model, (c) result for intensity model (grayscale) and (d) result for chrominance model (hue component of the HSV colour space).

The first scenario realizes home/surrounding protection against unauthorized intrusion. Fig. 5 contains a sample frame with the person walking in a garden and the foregrounds extracted for three various background models. It is noticeable that intensity (c) and chrominance (d) models are noisier than RGB model (b) and include more false detections. However, moving object was properly detected in each model.



Fig. 5 Experimental results for scenario A

Fig. 6 contains experimental results for two sample frames presenting a person, who enters the scene, falls after few seconds and lies still. The first row in Fig. 6 shows that RGB (b) and intensity (c) models are significantly influenced by the

shadow area, while in chrominance model (d) shadow is not detected. The second row in Fig 6 depicts the simulation of the lack of movement and its consequences: intensity (c) and chrominance (d) models incorporated object to the background. Moreover, other false detections appear on each resulted foreground image.

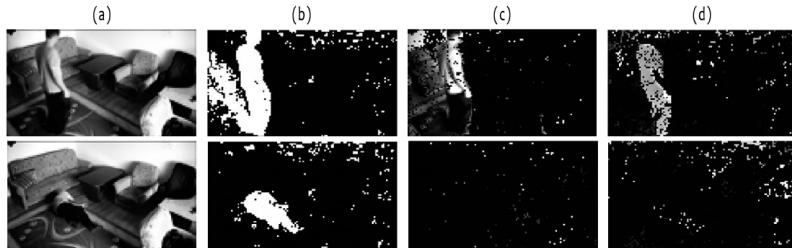


Fig. 6 Experimental results for scenario B — before the fall (first row) and after the fall (second row)

Fig. 7 contains a sample frame with a person with hands raised up and the extracted foregrounds. The results for different models vary significantly. Only in RGB model (b) the proper object's silhouette is detected, but at the same time the largest number of false detections is included. Intensity (c) and chrominance (d) models are less noisy, however it is difficult to recognize the moving object.

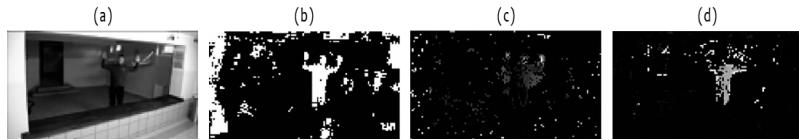


Fig. 7 Experimental results for scenario C

In conclusion, it is not easy to determine which of the implemented and tested background models is the best. There are some errors introduced on the object detection stage such as false detections caused by background noise, shadow areas or image compression. Additionally, there is a problem of correct objects elimination by models utilizing hue component (if there is no difference between chrominance points of the object and the background). Unfortunately, it is impossible to eliminate all aforementioned factors by a single background model. On the other hand, the utilization of all models simultaneously will be time consuming. Therefore, after the analysis of 'pros and cons', the application of RGB model was suggested. It introduced the least number of false detections and the existing noise level can be lowered by the appropriate post-processing. Moreover, some performance tests confirmed our choice. They were performed for RGB background model implemented in C with the use of *OpenCV* library. We measured the amount of time required to adapt pixel values on the basis of each single frame obtained from the video

sequence characteristic for scenario A (see Fig. 2 for sample frames), recorded at 30 fps. In most cases the amount of time needed for calculations was less than 0.01 of a second. Therefore, it is possible to assume that the system performance will be close to the real-time performance.

5 Conclusions

In this paper, we described the results of the experiments on various background models applied for the extraction of foreground regions in the process of background subtraction. The proposed solutions were examined and evaluated with the use of the SmartMonitor system test database in order to determine the best system working parameters for each scenario.

SmartMonitor is being currently developed as an innovative visual surveillance system based on video content analysis algorithms. It is intended to utilize only commonly available hardware and be affordable for individual users. The main advantage of the system will come from its customizability, i.e. the possibility of setting individual safety rules and adjusting the system sensitivity degrees.

The experiments shown that each model has its advantages and drawbacks, but the utilization of all models simultaneously will take too much time. Taking this into consideration, the RGB background model was selected. It is affected by the lowest level of noise and false detections.

Acknowledgements. The project '*Innovative security system based on image analysis — "SmartMonitor" prototype construction*' (original title: *Budowa prototypu innowacyjnego systemu bezpieczeństwa opartego o analizę obrazu — "SmartMonitor"*) is the project co-founded by European Union (project number PL: UDA-POIG.01.04.00-32-008/10-01, Value: 9.996.604 PLN, EU contribution: 5.848.800 PLN, realization period: 07.2011-04.2013). *European Funds — for the development of innovative economy (Fundusze Europejskie — dla rozwoju innowacyjnej gospodarki)*.

References

1. Forczmański, P., Seweryn, M.: Surveillance Video Stream Analysis Using Adaptive Background Model and Object Recognition. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part I. LNCS, vol. 6374, pp. 114–121. Springer, Heidelberg (2010)
2. Forczmański, P., Frejlichowski, D., Nowosielski, A., Hofman, R.: Current trends in developing of intelligent visual monitoring systems. Methods of Applied Computer Science (29), 19–32 (2011) (in Polish)
3. Frejlichowski, D.: Automatic localisation of moving vehicles in image sequences using morphological operations. In: Proceedings of the 1st IEEE International Conference on Information Technology, Gdansk, pp. 439–442 (2008)

4. Frejlichowski, D., Forczmański, P., Nowosielski, A., Gościewska, K., Hofman, R.: SmartMonitor: An Approach to Simple, Intelligent and Affordable Visual Surveillance System. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) IC-CVG 2012. LNCS, vol. 7594, pp. 726–734. Springer, Heidelberg (2012)
5. Frejlichowski, D., Gościewska, K., Forczmański, P., Nowosielski, A., Hofman, R.: SmartMonitor: recent progress in the development of an innovative visual surveillance system. *Journal of Theoretical and Applied Computer Science* 6(3), 28–35 (2012)
6. Gurwicz, Y., Yehezkel, R., Lachover, B.: Multiclass object classification for realtime video surveillance systems. *Pattern Recogn. Lett.* 32, 805–815 (2011)
7. Javed, O., Shafique, K., Shah, M.: A Hierarchical Approach to Robust Background Subtraction Using Color and Gradient Information. In: Workshop on Motion and Video Computing, pp. 22–27 (2002)
8. Kaewtrakulpong, P., Bowden, R.: An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection. In: Proceeding of the 2nd European Workshop on Advanced Video Based Surveillance Systems, Computer Vision and Distributed Processing, pp. 1–5. Kluwer Academic Publisher (2001)
9. Piccardi, M.: Background Subtraction Techniques: A Review. *IEEE International Conference on Systems, Man and Cybernetics* 4, 3099–3104 (2005)
10. Robertson, N., Reid, I.: A general method for human activity recognition in video. *Comput. Vis. Image Und.* 104, 232–248 (2006)
11. Sen-Ching, S.C.S., Kamath, C.: Robust Techniques for Background Subtraction in Urban Traffic Video. In: Bhaskaran V, Panchanathan S (eds) *Visual Communications and Image Processing*, vol. 5308, pp. 881–892 (2004)
12. Stauffer, C., Grimson, W.E.L.: Adaptive Background Mixture Models for Real-Time Tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246–252 (1999)
13. Wang, W., Chen, D., Gao, W., Yang, J.: Modeling Background from Compressed Video. *IEEE Transactions on Circuits and Systems for Video Technology* (5), 670–681 (2008)
14. Zivkovic, Z.: Improved Adaptive Gaussian Mixture Model for Background Subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition, vol. 2, pp. 28–31 (2004)
15. Bosch IVA 4.0 Commercial Brochure,
<http://resource.boschsecurity.com/>

Repeatability Measurements for 2D Interest Point Detectors on 3D Models

Simon R. Lang, Martin H. Luerssen, and David M.W. Powers

Abstract. Interest point detectors typically operate on 2D images, yet these frequently constitute projections of real 3D scenes [8]. Analysing and comparing the performance of these detectors as to their utility at tracking points in a 3D space is challenging. This paper demonstrates a virtual 3D environment which can measure the repeatability of detected interest points accurately and rapidly. Real-time 3D transform tools enable easy implementation of complex scene evaluations without the time-cost of a manual setup or mark-up. Nine detectors are tested and compared using evaluation and testing methods based on Schmid [16]. Each detector is tested on 34 textured and untextured models that are either scanned from physical objects or modelled by an artist. Rotation in the X, Y, and Z axis as well as scale transformations are tested on each model, with varying degrees of artificial noise applied. Results demonstrate the performance variability of different interest point detectors under different transformations and may assist researchers in deciding on the correct detector for their computer vision application.

1 Introduction

Interest points are pixel coordinates specifying an image location that contains a simple, localised, yet information-rich feature. These often serve as a means for tracking a scene's camera motion or an object's motion over time. In Schmid's classic work [16], the performance of interest point detectors was evaluated by establishing a homography between two interest points detected in the same scene at different perspectives, an approach that has since been used in a number of related

Simon R. Lang · Martin H. Luerssen · David M.W. Powers
School of Computer Science, Engineering and Mathematics
Flinders University, PO Box 2100,
Adelaide, South Australia 5001
e-mail: {simon.lang, martin.luerssen}@flinders.edu.au,
david.powers@flinders.edu.au

studies [4–6, 12]. This paper aims to revisit Schmid’s work and recreate the same experimental conditions in a much more versatile *virtual 3D* environment, which can utilise complex and precise transformations that would have been time consuming or impossible to perform in a real world environment. The position of interest points on 3D models can thus be quickly and accurately determined so that their repeatability—and by extension the performance of the detector that found these points—can be properly measured.

2 Background

Schmid’s [16] work evaluated the performance of interest point detectors in a scene using a *repeatability* measure. By testing the repeatability of interest points in two slightly different images I_1 and I_i , we can determine how well the detector can track these points across the applied transformation. Interest points x_1 can be detected in the image of a scene, I_1 . The scene is then changed by an affine transformation T_{1i} into a different scene and corresponding image, I_i , that has a different set of interest points x_i . Unlike Schmid, we operate on virtual 3D spaces, so T_{1i} as well as the projection transformations P_1 and P_i are precisely known. This allows a ray to be cast into the scene for each point in x_1 to determine, after intersection with the 3D model, which particular 3D point X in the scene corresponds to it; these points are subsequently also transformed by T_{1i} and subsequently projected into \tilde{x}_1 . The repeatability of each interest point in x_1 can then be gauged by whether there is an interest point in x_i around each transformed \tilde{x}_1 within a radial pixel distance threshold ε . The process is illustrated in figures 1 and 2 and can also be applied to x_i via an inverse T_{1i}^{-1} . Interest points must be located within the viewport after transformation, i.e., if points from \tilde{x}_1 appear in I_i and \tilde{x}_i points appear in I_1 then those

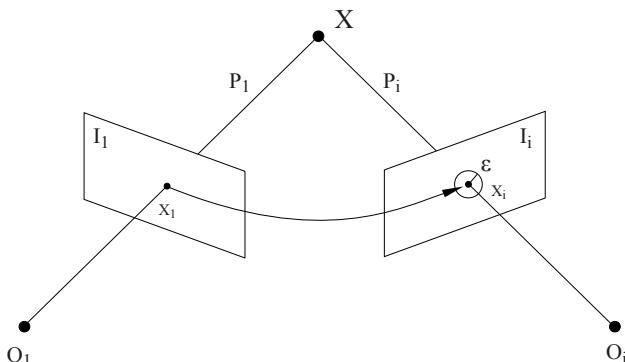


Fig. 1 Schmid repeatability: the points x_1 and x_i are the projections of a 3D point X onto images I_1 and I_i . A detected point x_1 is repeated if x_i is detected in the ε -neighbourhood of x_1 .

points are considered to share the same working area and are included for testing of repeatability. This determines the set of “repeated” points as:

$$R_i(\varepsilon) = \{(\tilde{x}_1, x_i) | dist(T_{1i}, \tilde{x}_1, x_i) < \varepsilon\}$$

A repeatability rate from 0 to 1.0 is derived from this as a ratio of repeated features divided by the lowest number of detected features between the image pairs. This is represented as:

$$r_i(\varepsilon) = \frac{|R_i(\varepsilon)|}{\min(n_1, n_i)}$$

where $n_1 = |\tilde{x}_1|$ and $n_i = |\tilde{x}_i|$ (visible subsets considered only).

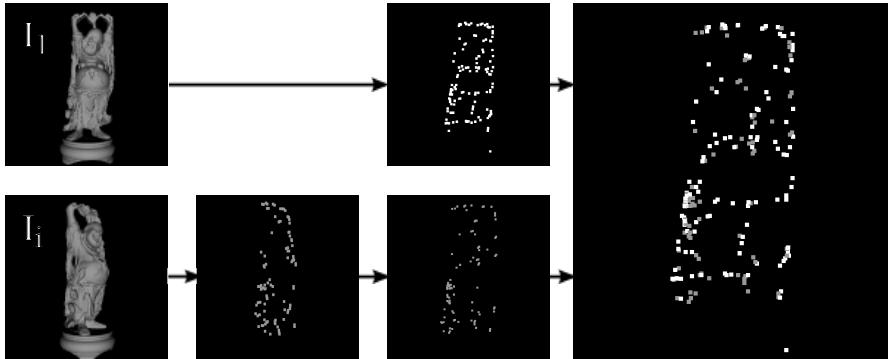


Fig. 2 Illustration of 3D model being rotated from 0° shown in I_1 , to 50° in I_i . I_1 and I_i are processed to produce interest points x_1 and x_i . 3D correspondences are determined via inverse projection and rotated by 50°, so that each point in x_i can be mapped to I_1 (and vice versa, not shown). Repeatability can now be measured easily by testing for points in close neighbourhood.

3 Method

A set of popular interest point detectors are tested, including Harris [7], KLT [2], FAST [14, 15], SIFT [11] and SURF [9]. The VIGRA [10] computer vision library additionally offers its own implementations of the Harris [7], Rohr [13], Foerstner [3] and Beaudet [1] detectors. The detectors’ parameter settings are left at their respective default values.

Models are manipulated within an OpenGL world space and rendered in a 400x400 pixel OpenGL viewport using linear filtering for textures. The models used are a diverse combination of 34 scanned and artist-made models. Figure 3 shows a small subset of these. Of these models, 15 are scanned untextured, 10 are artist-made, textured models, 8 are scanned textured models, and one is an unscanned,

untextured, artist-made model, which were all sourced from various research and model repository sites¹²³⁴⁵. Each model's initial rest position was oriented to help minimise occlusion and easily fit within the viewport. Interest points that exceed the viewport area at any stage before or after transformation are excluded from the repeatability score. Interest points that do not correspond to any feature of the model were also culled by establishing a bounding box around the model's dimensions. This predominantly includes interest points existing just beyond the edge of the model, which would inverse-project to the distant backplane of the 3D scene and therefore do not practically contribute to the repeatability score.



Fig. 3 Sample of models used in testing

The structural variability of the models necessitates that the robustness of each detector is tested on as many model positions as possible – an approach also followed by Schmid's original work. Since many applications for interest point detectors are applied to multiple, sequentially linked frames [8], it seems reasonable to develop a testing regimen that would fit this. Schmid's approach of limiting rotation to within -50° to 50° is retained for X and Y rotation since occlusion is much more likely at more extreme rotations, as well as rotating 180° face on in the Z axis. XY scale (i.e., zooming into the image) has been arbitrarily capped to a scaling factor of 4.0. Incrementation for X and Y rotation has been set at 5°, which means that the total 105° rotation is covered in 21 frames. Z rotation was made in 10° increments and ranged from 0° to 180°. The X, Y and Z rotations are all compared to the rest position at 0°. Scale in the XY axis starts at 1.0 in the rest position and is incrementally increased by 0.25 up to the aforementioned maximum.

The noise filter used for testing is a simple luminance based filter that randomly modifies the luminance of each pixel within a specified range (tolerance) of the total luminance space; a tolerance of 1.0 would produce completely random luminance values along a uniform distribution. Each of the repeatability evaluations were done using no noise, noise with a tolerance of 0.05 and noise with a tolerance of 0.10. The resultant image graininess assists to determine how reliable the feature detectors

¹ <http://users.cms.caltech.edu/keenan/models.html>

² <http://people.csail.mit.edu/tmertens/texttransfer/data/index.html>

³ <http://graphics.stanford.edu/data/3Dscanrep/>

⁴ <http://www.sci.utah.edu/wald/animrep/>

⁵ <http://www.turbosquid.com/3d>

are when the intensity of pixels are changed slightly and randomly, as happens in natural image capture. For simple illustration, Figure 4 demonstrates how the noise filter affects the luminance of pixels in the image.

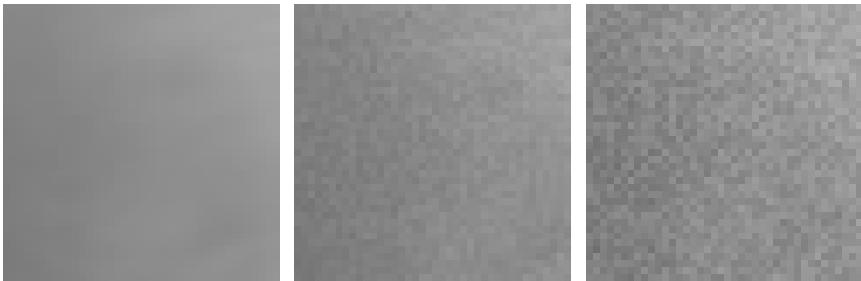


Fig. 4 Zoomed in sample of noise filter with no noise, noise with 0.05 tolerance and noise with 0.10 tolerance

4 Results

Interest point repeatability depends on the chosen ε value. Lower ε values are generally preferable, as they require interest points to lie very close together to be regarded as describing the same image feature, whereas larger values might lead to confusion between different features. However, the low value of 0.5ε , where two interest points correspond to each other only if they are closer than half a pixel in distance, yielded very poor performance in our tests, as shown in Figure 5 for each level of noise across all tested transformations. Further investigation revealed that this could be reproduced by simply changing the direction of the scanning subwindow of the detectors, so this level of inaccuracy appears to be surprisingly intrinsic and common to the detector implementations. Results at 1.5ε , i.e. a single pixel tolerance within the full Moore neighbourhood, may hence be more representative for purposes of comparing these detectors, and for this reason and for the sake of brevity, we will focus on 1.5ε when illustrating further outcomes.

Each of the transform tests in figures 6, 7, 8 and 9 illustrate the mean repeatability of each detector for all 34 models tested. Each level of noise tested shows that the performance of most detectors degrades as more noise is introduced. In some cases such as the Harris detector, the performance degradation is quite marked. In every test, Harris with default settings demonstrates very little robustness as more noise is introduced, which is observable even at the lower noise level, but can most easily be seen in figure 5c. Interestingly, the FAST detector has the best overall performance, and even performs unexpectedly better with a small amount of noise (see figure 5b), although it suffers a steep performance decline—particularly for small transformations—at the higher noise level tested here.

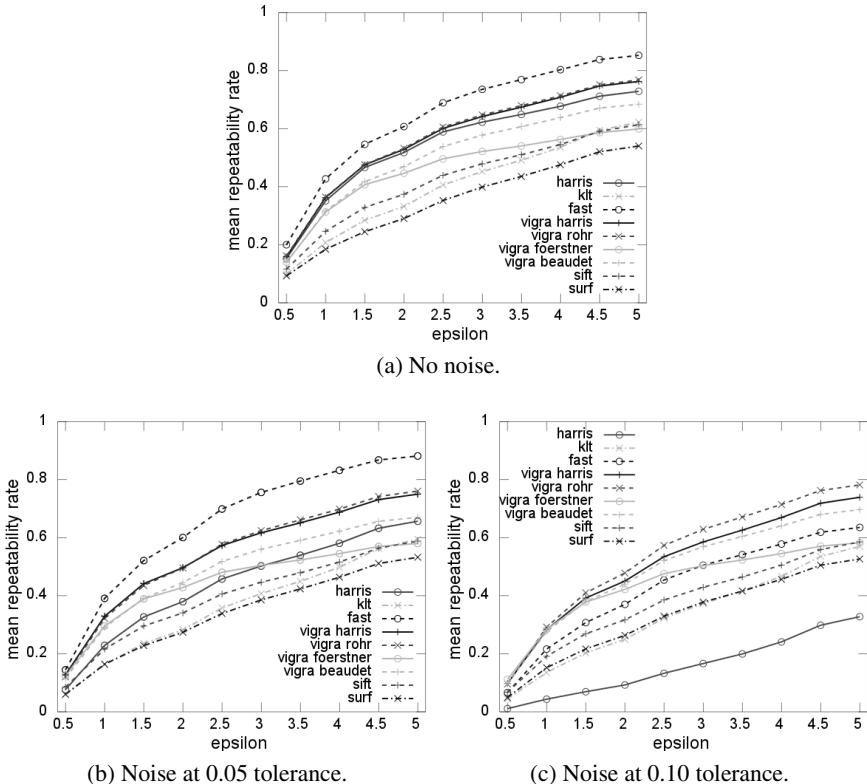
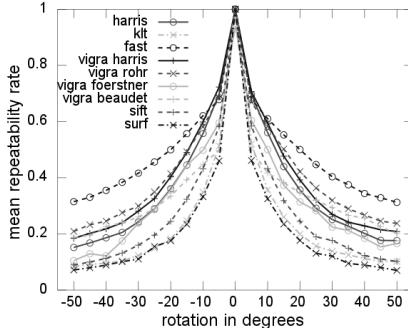


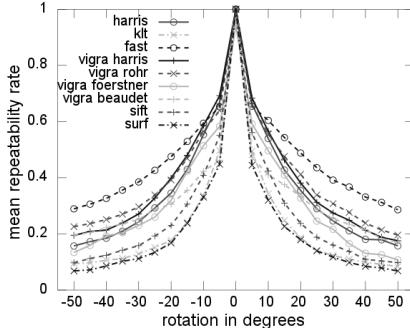
Fig. 5 Mean of 0.5ϵ to 5.0ϵ , for all tests, with all models

Overall, compared to the other interest point detectors, the FAST detector is most robust across larger transformations, as demonstrated in figures 6 to 9. Only at rotations of less than 10° , scale changes under 25%, and very high noise levels do the other detectors remain competitive. Conversely, SIFT, SURF, and KLT constitute the bottom of the pack, with weak capabilities in following features across larger model rotations and scaling. However, it is worth remembering that this assessment is only based on a single performance measure—repeatability—and there are other properties of these detectors that can make them desirable for a particular application. The VIGRA-included detectors performed within a comparatively similar mid-range performance band, and there are few surprises to be noted here. As with all other detectors—although only weakly observed with SIFT and SURF—they exhibit a particular sensitivity to multiples of 90° rotations along the Z-axis.

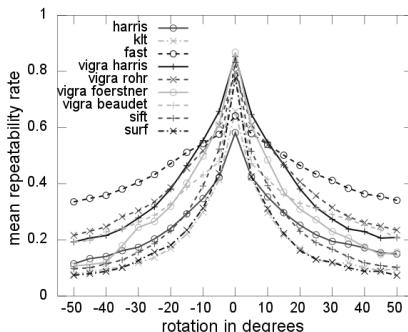
Though some effort was taken to reduce the effect of point occlusion on repeatability, the models themselves may self-occlude features when rotated along the X and Y axes. Conventionally, points occluded this way should be considered outside of the working area, but such situations are not fully handled by our system. At



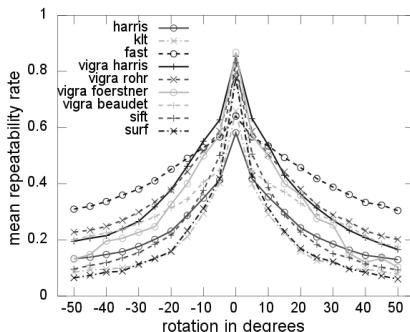
(a) No noise.



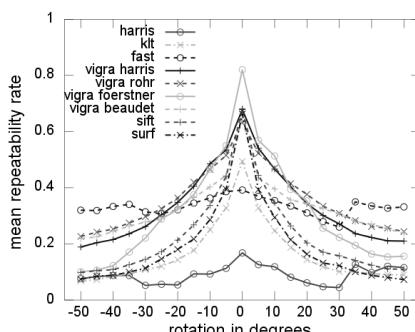
(a) No noise.



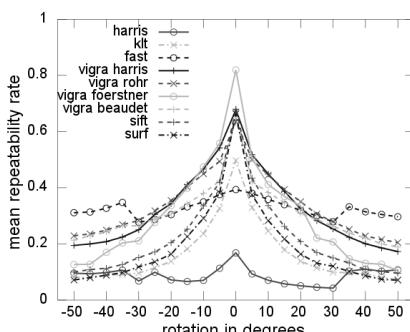
(b) Noise at 0.05 tolerance.



(b) Noise at 0.05 tolerance.



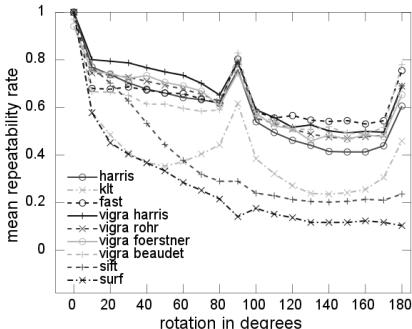
(c) Noise at 0.10 tolerance.



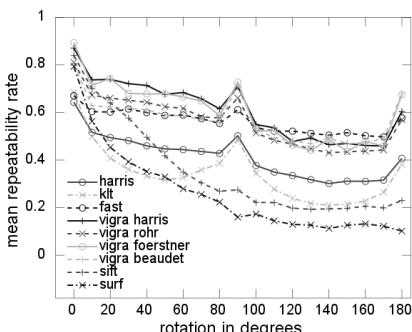
(c) Noise at 0.10 tolerance.

Fig. 6 X rotation at 1.5ϵ with all models**Fig. 7** Y rotation at 1.5ϵ with all models

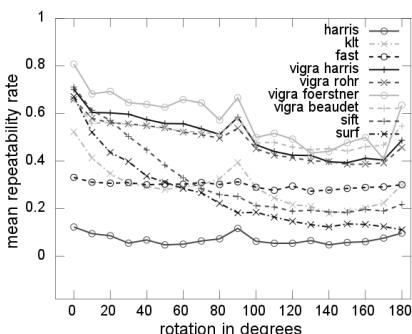
this stage, we in fact do not know how much of a significant factor this is as far as repeatability is concerned, although large rotations would obviously result in extensive occlusion. Because depth does not play a factor in measuring the ϵ distance of points in images I_1 and I_i , we can find situations where interest points at a greater relative Z depth and closer relative to the XY axes can be considered closer than



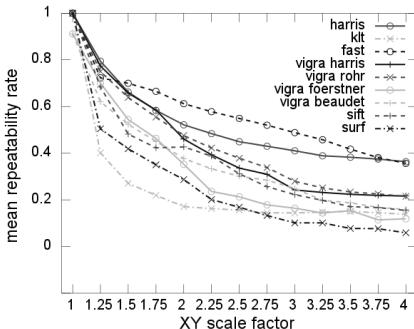
(a) No noise.



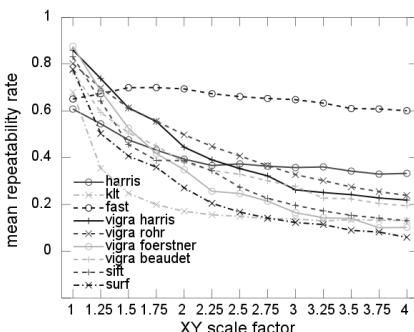
(b) Noise at 0.05 tolerance.



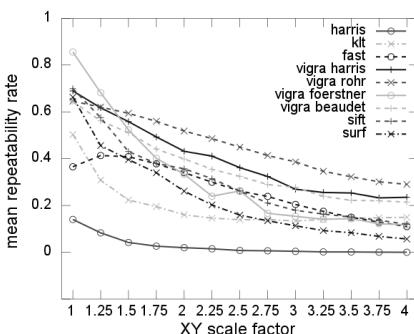
(c) Noise at 0.10 tolerance.

Fig. 8 Z rotation at 1.5ϵ with all models

(a) No noise.



(b) Noise at 0.05 tolerance.



(c) Noise at 0.10 tolerance.

Fig. 9 Scale XY rotation at 1.5ϵ with all models

interest points that share the same Z depth yet are measured as further away from the reference interest point in image I_1 . However, situations like this are only a problem if the point from I_i is also an occluded point in I_1 . A larger impact on repeatability is expected from those interest points that arise along the edge of the model

(bordering the scene background) but which do not represent an actual feature on the model itself. These do not consistently transform with the rest of the model, e.g., a rotation of a sphere will not change anything about the position of the sphere's edge. Currently, such interest points are not filtered out (except if they lie outside the actual model), as they are very distinct features of the image and obvious targets for image point detectors, but more careful consideration of their impact is needed.

5 Conclusion

The feature tracking performance of nine low-level interest point detectors was systematically evaluated on Schmid's repeatability measure [16] implemented within a 3D OpenGL virtual environment. Rotations in the X, Y, and Z axis as well as scaling was applied to 34 3D models, including scanned and artist-made objects, and tested with varying degrees of artificial noise. The demonstrated system was capable of finely differentiating interest point repeatability and robustness in a precise and highly reproducible way, albeit at the cost of offering only a synthetic benchmark. The results indicate a general, but not universal, superiority of the FAST detector, and should offer a useful insight into the behaviour of these interest point detectors for researchers intending to utilize them in their computer vision tasks. In future, the proposed framework is intended to be used in rapidly and potentially automatically prototyping new detector types for specific applications, a process in which synthetic testing such as employed here is highly practical.

Acknowledgements. This work is partially supported by the Chinese Natural Science Foundation under Grants (No.61070117) and the Beijing Natural Science Foundation under Grant (No.4122004).

References

1. Beaudet, P.: Rotationally invariant image operators. In: Proc. Intl. Joint Conf. on Pattern Recognition, pp. 579–583 (1978)
2. Carlo Tomasi, T.K.: Detection and tracking of point features. In: Carnegie Mellon University Tech. Rpt (1991)
3. Förstner, W.: A feature based correspondence algorithms for image matching. Intl. Arch. Photogrammetry and Remote Sensing 24, 160–166 (1986)
4. Gauglitz, S., Hllerer, T., Turk, M.: Evaluation of interest point detectors and feature descriptors for visual tracking. Int. Journal of Comp. Vis. 94, 335–360 (2011)
5. Gil, A., Mozos, O., Ballesta, M., Reinoso, O.: A comparative evaluation of interest point detectors and local descriptors for visual slam. Machine Vision and Applications 21, 905–920 (2010)
6. Guillaume Gals, S.C., Crouzil, A.: Complementarity of feature point detectors. Intl. Joint Conf. on Comp. Vis. Theory and App. (2010)
7. Harris, C., Stephens, M.: A combined corner and edge detector (1988)

8. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York (2003)
9. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006, Part I*. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
10. Köthe, U.: Generische programmierung für die bildverarbeitung. PhD Thesis, Universität Hamburg (2000)
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Intl. Journal of Comp. Vis.* 60, 91–110 (2004)
12. Olague, G., Trujillo, L.: Evolutionary computer assisted design of image operators that detect interest points using genetic programming. *Image and Vision Computing* 29, 484–498 (2011)
13. Rohr, K.: Modelling and identification of characteristic intensity variations. *Image and Vis. Comp.* 10, 66–76 (1992)
14. Rosten, E., Drummond, T.: Fusing points and lines for high performance tracking. In: *IEEE Intl. Conf. on Comp. Vis.*, vol. 2, pp. 1508–1511 (October 2005)
15. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: *European Conf. on Comp. Vis.*, vol. 1, pp. 430–443 (May 2006)
16. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. *Intl. Journal of Comp. Vis.* 37, 151–172 (2000)

Extended Investigations on Skeleton Graph Matching for Object Recognition

Jens Hedrich, Cong Yang, Christian Feinen, Simone Schäfer,
Dietrich Paulus, and Marcin Grzegorzek

Abstract. Shape similarity estimation of objects is a key component in many computer vision systems. In order to compare two shapes, salient features of a query and target shape are selected and compared with each other, based on a predefined similarity measure. The challenge is to find a meaningful similarity measure that captures most of the original shape properties. One well performing approach called Path Similarity Skeleton Graph Matching has been introduced by Bai and Latecki. Their idea is to represent and match the objects shape by its interior through geodesic paths between skeleton end nodes. Thus it is enabled to robustly match deformable objects. However, insight knowledge about how a similarity measure works is of great importance to understand the matching procedure. In this paper we experimentally evaluate our reimplementation of the Path Similarity Skeleton Graph Matching Algorithm on three 2D shape databases. Furthermore, we outline in detail the strengths and limitations of the described methods. Additionally, we explain how the limitations of the existing algorithm can be overcome.

Keywords: Skeleton, Skeleton Graph, Graph Matching, Shape Recognition.

1 Introduction

Sensory devices have become increasingly affordable. The processing power as well as storage space have been drastically improved in last decades. The amount of image data is growing rapidly. On the one hand, recording and consumption of such data has been getting easier. On the other hand, complexity of searching and reasoning complicates the access to data [17]. Compared to well-known data types like

Jens Hedrich · Simone Schäfer · Dietrich Paulus
University of Koblenz-Landau, Germany

Cong Yang · Christian Feinen · Marcin Grzegorzek
University of Siegen, Germany

plain text documents, images are much more sophisticated to manage. For example, searching in semi-structured data like text documents are less complex since searching conditions can be controlled by syntactic means. However, queries based on images have to be implemented by semantic aspects which are not explicitly known previously. Thus, similarity measures are an ongoing research topic, which is an important contribution to various applications, like multimedia retrieval and object recognition [12]. In order to compare two shapes, salient features of both have to be extracted and compared with each other. Afterwards, shape features of the query object are compared to those of the target. Therefore, an appropriate and predefined similarity measure has to be selected. One challenge is to find a meaningful similarity measure that captures most of the original object's properties. Unlike other similarity measures, the proposed method in this paper establishes correspondences between *skeleton and nodes*. The basic concept is proposed in [2], including a comprehensive evaluation which shows promising results for comparing 2D objects based on skeletons.

We start by discussing the related work and providing a brief explanation of the Path Similarity Skeleton Graph Matching algorithm proposed by Bai et al. [2] (Section 2). In Section 3, we outline some limitations and issues which can appear within the matching process. Section 4 presents the recognition performance of the reimplemented algorithm on three 2D shape databases. Finally, we conclude our work in Section 5.

2 Related Work

Algorithms to analyse objects by shape can be typically categorized by the data representation, namely (i) point set representation, (ii) boundary representation, and (iii) medial representation. The *point set representation* is an unorganized point set. One-to-one correspondences between two point sets are established based on meaningful descriptors [4]. The goal is to find corresponding pairs of points in both shapes that have the highest similarity. The *boundary representations* represent an object by its hull; e.g. *snakes* [7] are used to match objects. Thereby, one idea is to measure the matching energy which is needed to match two contours [18]. The *medial representation* describes the object by its interior, e.g. *skeletons* (also called *medial axes* or *symmetry axes*) are the most propagated medial object representations. They include essential topology and geometrical information [10]. In comparison to boundary representation approaches, skeleton-based methods show their advantages in matching of deformable objects. They are more robust to overlaps, deformations or misplaced object parts [14]. However, skeletons are sensitive to noisy input data, which increases the complexity of the skeletal structure and subsequently the matching complexity. Therefore, the skeleton's quality depends on pre- and post-processing methods, e.g., contour sampling or skeleton pruning. The advantage is that the matching can be reduced to the graph matching problem. Matching single salient skeleton points, e.g. junction nodes or end nodes, is a

further approach. In [8, 9] shock trees – a variant of skeletons – are used for shape-comparison based on an edit-distance algorithm. The edit distance is computed by traversing the rooted shock tree while edit operations are being applied to the traversed edges. All edit operations are associated to a predefined edit cost. The idea is to transform one skeleton branch into an opponent branch. Thus, the similarity between two shapes is measured by summing up the cost of deformations between the shock transitions. Many matching approaches enforce one-to-one correspondences, but noisy image data deal with the problem that one-to-one matchings are not always possible. In [5, 6] the authors present a technique for many-to-many matching of medial axis graphs. The idea is to embed the nodes of two graphs into a fixed-dimension Euclidian space and using the Earth Mover’s Distance to enable a many-to-many matching between nodes of the graphs.

2.1 Path Similarity Skeleton Graph Matching

Following contents provide a description of the Path Similarity Skeleton Graph Matching algorithm, proposed by Bai and Latecki in [2]. Our aim is to provide a deep understanding of the concept in order to explain the investigated limitations of this algorithm in Section 3. The algorithm can be divided into three major parts: (i) getting a compact skeleton representation, (ii) computing the matching costs between the end points, (iii) repeating the latter procedure for all combinations of end nodes. The final matching of the skeleton end nodes is performed by applying the Hungarian algorithm.

2.1.1 Skeleton Representation

A key concept is the use of information about skeleton paths. The skeleton representation incorporates the main contour information. The example of the two bird shapes in Fig. 1a-b will guide through the explanations. A skeleton path $p(v_m, v_n)$ is defined as the shortest path between a pair of end nodes v_m and v_n passing the skeletal structure (see Fig. 1c). The skeleton paths are found by constructing a weighted skeleton graph. The edge weight is defined by the length of the corresponding skeleton branch. Hence, it is possible to apply a shortest path algorithm (e.g. Dijkstra’s algorithm). Additionally, information about the object contour is included. Thus, a skeleton path $p(v_m, v_n)$ is sampled with M equidistant points. Afterwards, all radii of maximal disks, measured at each sampled point t , are noted within a path vector (cf. Eq. 1 and Fig. 1d):

$$R_{m,n} = (R_{m,n}(t))_{t=1,2,\dots,M} = (r_1, r_2, \dots, r_M), \quad (1)$$

In [2] the distance of each sampled skeleton point t to its feature point is approximated by a distance transform $DT(t)$, in our case a Euclidian distance map is computed. Afterwards, the distance is normalized to make the method invariant to scale. Finally, the distances are approximated and normalized as follows:

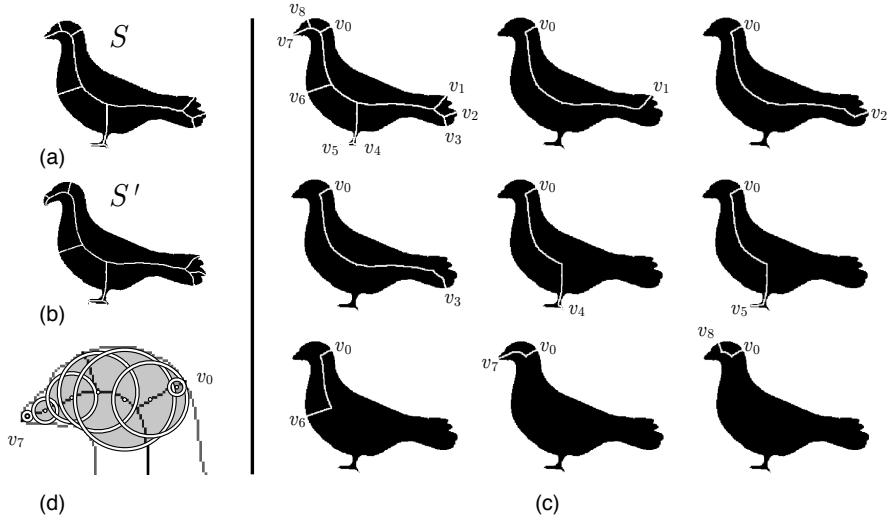


Fig. 1 (a-b) Example for two skeletons S and S' to be matched. (c) The complete skeleton (top left) and all skeleton paths emanating of one example end node. (d) Sampling of a skeleton path. The sampling points are indicated with white dots. The distance to their feature points is indicated by the black circles. For the skeleton path representation, the normalized distance of the skeleton points to their feature points is measured and noted in the skeleton path vector.

$$R_{m,n} = \frac{DT(t)}{\frac{1}{N_0} \sum_{i=1}^{N_0} DT(s_i)}, \quad (2)$$

where N_0 is the number of pixels in the original shape and $s_i (i = 1, 2, \dots, N_0)$ varies over all N_0 pixels within the shape. The ordered list of M distance values is obtained for each skeleton path. All distance values are noted in the path vector (cf. Eq. 1).

2.1.2 Dissimilarity between End Nodes

In order to compute the matching costs for two end nodes, a similarity measure is necessary. Therefore, the dissimilarity of two skeleton paths is given by the *path distance* shown in (Eq. 3) (where r_i and r'_i are radii of maximal disks of the path vectors R and R' , l and l' are the length of the skeleton paths $p(u, v)$ and $p(u', v')$). The influence of the path length is weighted by the factor $\alpha \in \mathbb{R}^+$. To make the approach invariant to scale, the lengths are normalized. The assumption is that similar skeleton paths have consecutive skeleton points with similar radii of maximal, inscribed discs.

$$pd(p(u, v), p(u', v')) = \sum_{i=1}^M \frac{(r_i - r'_i)^2}{r_i + r'_i} + \alpha \frac{(l - l')^2}{l + l'} \quad (3)$$

All path distances for one pair of end nodes are combined in one *path distance matrix* (**PDM**) (Eq. 4). Two skeleton graphs G with $K + 1$ end nodes and G' with $N + 1$ end nodes are matched with $K \leq N$. The end nodes v_i and v'_j of G and G' are ordered by traversing the object contours in clockwise direction. This leads to an ordered list of end nodes: $\{v_{i0}, v_{i1}, \dots, v_{iK}\}$ for G and $\{v'_{j0}, v'_{j1}, \dots, v'_{jN}\}$ for G' .

$$\mathbf{PDM}(v_{i0}, v'_{j0}) = \begin{pmatrix} pd(p(v_{i0}, v_{i1}), p(v'_{j0}, v'_{j1})) & \cdots & pd(p(v_{i0}, v_{i1}), p(v'_{j0}, v'_{jN})) \\ \vdots & \vdots & \vdots \\ pd(p(v_{i0}, v_{iK}), p(v'_{j0}, v'_{j1})) & \cdots & pd(p(v_{i0}, v_{iK}), p(v'_{j0}, v'_{jN})) \end{pmatrix} \quad (4)$$

To estimate a dissimilarity value for a pair of end nodes v_i and v'_j , the **PDM** is applied to the *Optimal Subsequence Bisection (OSB)* [11]. By using this approach, the problem of estimating the similarity of two end nodes is reduced to elastic matching of time series. One main advantage of using *OSB* is that outliers within the path distance matrix can be easily skipped. The **PDM** is computed for every combination of end nodes in two skeleton graphs and afterwards applied to the *OSB* function: $c(v_i, v'_j) = \text{OSB}(\mathbf{PDM}(v_i, v'_j))$. Subsequently, the resulting cost matrix C (cf. eq. 5) is used as input for the Hungarian algorithm. Hence, the matching problem is reduced to the classic assignment problem in a bipartite graph.

$$C(G, G') = \begin{pmatrix} c(v_0, v'_0) & c(v_0, v'_1) & \cdots & c(v_0, v'_N) \\ c(v_1, v'_0) & c(v_1, v'_1) & \cdots & c(v_1, v'_N) \\ \vdots & \vdots & & \vdots \\ c(v_K, v'_0) & c(v_K, v'_1) & \cdots & c(v_K, v'_N) \end{pmatrix} \quad (5)$$

3 Investigations on the Path Similarity Skeleton Graph Matching Algorithm

The following investigations are based on a reimplementation of the Path Similarity Skeleton Graph Matching algorithm proposed in [2]. Deduced from our experience, we detected three major limitations, which can occur in special matching cases: flipped images, 1-to-1 matching of the end nodes, and spurious skeleton branches.

3.1 Flipped Images

The cheapest path through a given path distance matrix (**PDM**) is estimated by the *OSB*-Function. It is assumed that the cheapest path for two corresponding end nodes goes from the upper left corner to the lower right corner. By traversing the matrix with the *OSB* function, it is not allowed to go backwards, neither in the rows nor in the columns. In the case of matching two similar shapes which head towards different directions, the correct matching costs cannot be estimated with the *OSB*

function (cf. Fig. 2a). In these cases, the actual cheapest path through the matrix is flipped and goes from the upper right corner to the lower left corner (cf. Eq. 6). This means that the *OSB* function is not able to estimate a reliable indicator for the similarity.

$$\text{PDM}(v_{i0}, v'_{j0}) = \begin{pmatrix} \dots & 7 & 5 & 8 & 0 \\ \dots & 2 & 3 & 0 & 8 \\ \dots & 5 & 0 & 3 & 5 \\ \dots & 0 & 5 & 2 & 7 \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \quad (6)$$

As a solution, we apply the *OSB* function twice: once for the original image, and once with one image flipped horizontally. From the resulting two match lists, the one with lower matching costs will be chosen as the real matching. In the most cases this works quite well. The minimal matching cost value is taken as similarity value. However, this strategy can fail in the case of strong dissimilarities between two shapes. Only, a more complex strategy which compares contour partitions in detail can overcome this failure. For example, the shapes in Fig. 2a are oriented in opposite directions, but the matching costs for the second run with one of the images flipped leads to lower matching costs than the first run. Thus, in the algorithm the two shapes are assumed to be oriented in the same direction, which in the end leads to an unsatisfying matching.

3.2 1-to-1 Matching

1-to-1 matching of end nodes is another issue which we identified. It is not always possible to assign a correct matching partner to each end node. For example Fig. 2b depicts an acceptable matching, but both skeletons have one additional end node that has no matching partner. Within the Hungarian algorithm each end node has to be assigned to one partner, even though they do not correspond. This is not only a limitation of this particular algorithm, but a problem of all matching algorithms that reduce the matching problem to a 1-to-1 matching in a bipartite graph. Using a different approach could be a solution to this problem. For example, the Earth

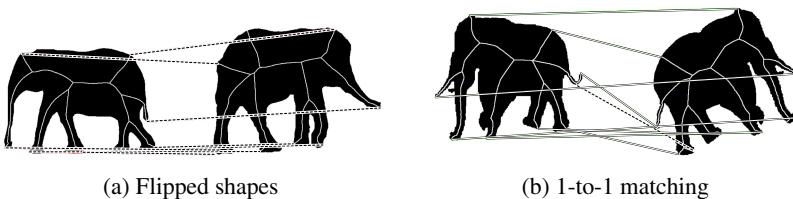


Fig. 2 a) If one shape is flipped, it is likely that the method fails to determine. b) 1-to-1 matching is not always possible. In this example, all matchings have been found correctly, but the two remaining end nodes with no matching partner in the other skeleton are matched.

Mover's Distance (EMD) as used in [13] also allows partial matchings. This could be a solution to better deal with noisy skeleton data.

3.3 Spurious Skeleton Branches

The Path Similarity Skeleton Graph Matching algorithm requires perfectly regularized skeletons, where each skeleton branch represents a significant visual part of the shape. Since each end node has to be assigned to an 1-to-1 matching partner, spurious skeleton branches have profound negative impact on the matching result. Fig. 3a illustrates wrong assignments between two elephant shapes, which are caused by a spurious branch in the tail of the left elephant (see Fig. 3b). After manually removing this branch the number of correct correspondences has significantly increased (cf. Fig. 3c). Thus, one has to make sure that the input skeletons do not contain spurious branches.

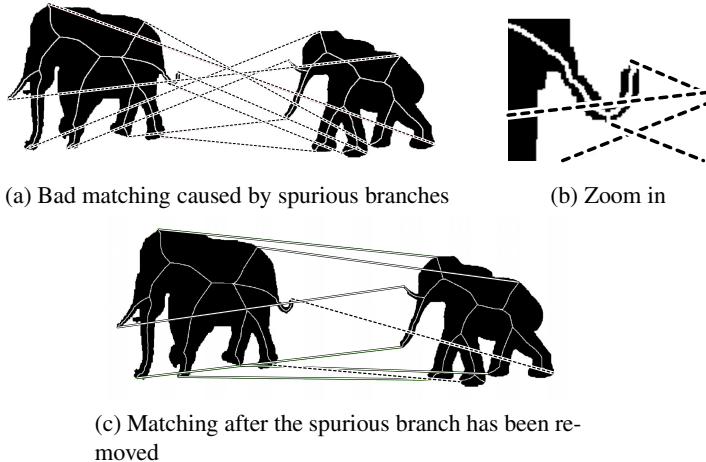


Fig. 3 One spurious branch can have a high impact on the matching result

4 Recognition Performance

Based on our reimplementation, we used several shape databases like Aslan and Tari [1], kimia-99 [15] and kimia-216 [16] to evaluate the recognition performance. The skeleton of each shape is computed by the *Discrete Curve Evolution* (DCE) algorithm [3]¹, with the parameters $\rho = 4$, $T_1 = 1$ and *number_vertex* = 15. The parameters for the similarity measurement were $M = 50$ and $\alpha = 40$. All shapes

¹ Available at <https://sites.google.com/site/xiangbai/BaiSkeletonPruningDCE.zip>

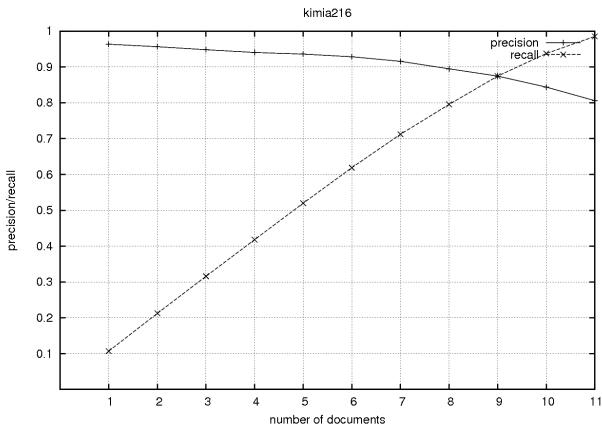


Fig. 4 Average precision and recall development in the kimia216 database with increasing number of result documents

from the databases has been used as a query. In order to rate the retrieval, the average precision is computed and for each query the maximal number of shapes within the retrieval class is returned. For each query on the Aslan and Tari database, three shapes are returned, which leads to a average precision of 0.93. For each query on the kimia-99 database 9 shapes are returned, which results in an average precision of 0.84. In the kimia-216 database 11 shapes are returned for each query, the average precision is 0.81 (cf. Fig. 4). Table 1 summarizes the number of all correct shapes for the first eleven retrieval results from the kima216 database in comparison to the values listed in the original paper. Obviously, the results are not as good as in the original paper. It is assumed that the input skeletons play a significant role. Several skeletons used in the experiments contain spurious branches, which had a profound impact on the matching quality and led to distorted overall similarity values. This effect has been observed in many of the query results. However, the originally used parameters for the DCE algorithm are not reported.

Table 1 Summary of correct shapes in the 1st, 2nd,... retrieval result

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th
Original paper	216	216	215	216	213	210	210	207	205	191	177
Our results	210	208	203	202	200	192	186	167	161	130	96

To verify this assumption, further experiments were performed on the more problematic classes of the Kima-216 database. This time, the skeletons in the database were pruned manually so that each skeleton branch represents a significant visual part of the original shape. As the significant parts of shapes of the same class should be quite similar, the skeletons get more comparable. Using the manually pruned

skeletons leads to better results in the performed queries. For example, the average precision for the queries from the 'bird'-class went up from 0.69 to 0.81, the average precision for the queries from the 'camel' class went up from 0.63 to 0.73 (see table 2). Additionally, it can be observed that the average precision value for the Kimia-99 and Kimia-216 database is worse than for the Aslan and Tari database. The reason for this partly lies in the composition of data in both databases. In the Aslan and Tari database, the algorithm's performance for non-rigid shapes is mainly evaluated. Parts of the shapes are bent and the shapes are similar to each other within a class. The main challenge in the Kimia-99 database is that several shapes are occluded.

Table 2 Example queries on the Kimia-216 database. In the left column, the query shape is shown. From left to right, the eight most similar shapes in the database are shown. The similarity to the query drops from left to right.

Query	1st	2nd	3rd	4th	5th	6th	7th	8th

5 Conclusion

In this paper we reimplemented the Path Similarity Skeleton Graph Matching algorithm [2] and performed a reevaluation on three shape databases (Aslan and Tari, kimia-99 and kimia-216). Additionally, we reported the limitations of the algorithm in detail for the first time. A fundamental understanding is necessary to understand upcoming issues during a matching process. The algorithm showed its advantages when dealing with non-rigid objects and articulated joints. An average precision of 0.93 (0.98 with manually pruned skeletons, respectively) for the Aslan and Tari database shows that shape deformations do not affect the skeleton's topology and the use of path radii has no impact on the matching results. The experiments with the Kimia-99 and Kimia-216 database also showed acceptable results with an average precision of 0.84 and 0.81, respectively. In this paper, we were not able to reproduce the excellent recognition results of the original paper, due to the lack of the not reported parameters for the skeletonization method. However, problems in the recognition performance occurred when shapes of different classes were similar. In addition, overlaps have a negative impact on the recognition results. A severe limitation of the algorithm is its requirement of optimal skeletons. The experiments showed that spurious branches in one of the skeletons lead to distorted matching

results in several cases. Consequently, this affects the object recognition performance when using the described matching algorithm in the retrieval system. In the future, we will investigate strategies how to deal with flipped images and pruning algorithms for skeletons to reduce the number of spurious branches. Furthermore, this matching algorithm could be used as an initial similarity measurement in a hierarchical and more complex object recognition system. Another challenge is the rapidly growing amount of 3D data. An extension of the Path Similarity Skeleton Graph Matching method towards the third dimension would be a great contribution to the object recognition community.

References

1. Aslan, C., Tari, S.: An axis-based representation for recognition. In: Proceedings of the Tenth IEEE International Conference on Computer Vision, pp. 1339–1346 (2005)
2. Bai, X., Latecki, L.J.: Path similarity skeleton graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(7), 1282–1292 (2008)
3. Bai, X., Latecki, L.J., Liu, W.-Y.: Skeleton pruning by contour partitioning with discrete curve evolution. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 449–462 (2007)
4. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(4), 509–522 (2002)
5. Fatih Demirci, M., Shokoufandeh, A., Keselman, Y., Bretzner, L., Dickinson, S.J.: Object recognition as many-to-many feature matching. *Int. J. Comput. Vision* 69, 203–222 (2006)
6. Fatih Demirci, M., Shokoufandeh, A., Dickinson, S.: Skeletal shape abstraction from examples. *IEEE Trans. Pattern Anal. Mach. Intelligence* 31(5), 944–952 (2009)
7. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
8. Klein, P., Srikantha, T., Sharvit, D., Kimia, B.: A tree-edit-distance algorithm for comparing simple, closed shapes. In: Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 696–704. Society for Industrial and Applied Mathematic, Philadelphia (2000)
9. Klein, P.N., Sebastian, T.B., Kimia, B.B.: Shape matching using edit-distance: an implementation. In: Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 781–790. Society for Industrial and Applied Mathematic, Philadelphia (2001)
10. Komala Lakshmi, J., Punithavalli, M.: A survey on skeletons in digital image processing. In: Proceedings of the International Conference on Digital Image Processing, pp. 260–269 (2009)
11. Latecki, L.J., Wang, Q., Koknar-Tezel, S., Megalooikonomou, V.: Optimal subsequence bijection. In: ICDM 2007: Proceedings of the 2007 Seventh IEEE International Conference on Data Mining, pp. 565–570. IEEE Computer Society, Washington, DC (2007)
12. Tadeusiewicz, R.: How Intelligent Should Be System for Image Analysis? In: Kwasnicka, H., Jain, L.C. (eds.) *Innovations in Intelligent Image Analysis*. SCI, pp. VX. Springer, Heidelberg (2011), <http://www.springer.com>
13. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vision* 40, 99–121 (2000)

14. Sebastian, T.B., Kimia, B.B.: Curves vs. skeletons in object recognition. *Signal Processing* 85(2), 247–263 (2005)
15. Sebastian, T.B., Klein, P.N., Kimia, B.B.: Recognition of shapes by editing shock graphs. In: *IEEE International Conference on Computer Vision*, pp. 755–762 (2001)
16. Sebastian, T.B., Klein, P.N., Kimia, B.B.: Recognition of shapes by editing their shock graphs. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 26(5), 550–571 (2004)
17. Tadeusiewicz, R.: What does it means "automatic understanding of the images"? In: *Proceedings of the 2007 IEEE International Workshop on Imaging Systems and Techniques*, pp. 1–3 (May 2007)
18. Younes, L.: Computable elastic distances between shapes. *SIAM J. Appl. Math.* 58, 565–586 (1998)

Low-Level Image Features for Stamps Detection and Classification

Paweł Forczmański and Andrzej Markiewicz

Abstract. This paper presents a novel approach to detect and classify stamp instances in scanned documents. It incorporates several methods from the field of image processing, pattern recognition as well as some heuristic. At first, color separation is applied in order to find potential stamps. Next, several methods aimed at detecting objects of specific shapes are employed. Then, isolated objects are extracted and classified using a set of shape descriptors. Selected features are rotation, scale and translation invariant, hence this approach does not depend on the document size and orientation. The experiments performed on a large set of real documents retrieved from the Internet gave encouraging results.

1 Introduction

Document analysis is one of the most important areas of digital image processing and pattern recognition. The task of detection and recognition of stamps (seals) on digital images is a very important problem. Since we are facing an important change in the technology - conversion from paper documents into digital ones - the need for algorithms and approaches having the ability to automatically segment and extract important document's elements is very high. Despite significant progress in this technology, this problem still remains open [8, 12]. The difficulty of stamp detection and classification is related mainly to the lack of templates. In general, there are no good templates for stamps. It is because there is no standard and commonly used stamp's representation. Stamps are complex objects, containing graphical and textual elements that can be placed on any position in the document. Stamps diversity

Paweł Forczmański · Andrzej Markiewicz
West Pomeranian University of Technology, Szczecin,
Faculty of Computer Science and Information Technology,
Żołnierska Str. 52, 71–210 Szczecin, Poland
e-mail: {pforczmanski, amarkiewicz}@wi.zut.edu.pl

comes from their orientation, shape, color, fonts, ornaments and quality of imprinting. Even two imprints of the same physical stamp can look very different. This same applies to other objects that were described in [4] or [9].

Recently, several methods oriented at automatic stamp detection and classification have been proposed. They work on color objects [8] or detect stamps of particular shape [12]. Unfortunately, there is no approach that detects and classifies all the diversity of stamps. There are also several approaches oriented at logo retrieval, which is a similar problem [13]. The newest ones include application of one of the several modern shape descriptors, i.e. SIFT/SURF [6, 7] or ART [5]. It should be however noted, that such approaches can not be directly employed to stamp detection. Although, there are many common geometrical features that characterize both objects [4, 13], the process of imprinting introduces strong distortions (noise, inconsistency, gaps, stains, etc.) in the stamp representation. Because of this fact and the diversity of stamps, even within a single class (in terms of shape), the application of such descriptors is not possible.

In this paper we present a new approach to detect stamps of different shapes and colors and extract them properly even if they are overlapped with signature or text. Because of the high variance in shape domain we focus on official stamps (with well-defined shape) of any particular color. In opposition to [8], the detection of black stamps is also possible. Shapes being detected are not limited to ovals and squares, like in [12]. Good results have been achieved detecting stamps in documents containing other similar objects such as logos and texts.

2 Algorithm Overview

The input for the proposed approach is given as a scanned or photographed document. It is important that the imaging plane is parallel to the document plane in order to preserve geometrical features of all elements. From the practical point of view, it is also crucial, that the image has quite high spatial resolution and is stored in a format that provides minimal loss of quality. The algorithm is developed to work on both color and monochrome images that are not subjected to any brightness equalization.

The output provides information about the number of stamps detected, their shapes, colors and coordinates.

The algorithm (see Fig. 1) consists of the following steps:

1. Color segmentation and candidates extraction (for color images),
2. Shape-specific detection:
 - detection of circles,
 - detection of lines,
 - detection of other shapes using heuristics,
3. Stamps verification,
4. Classification and final output.

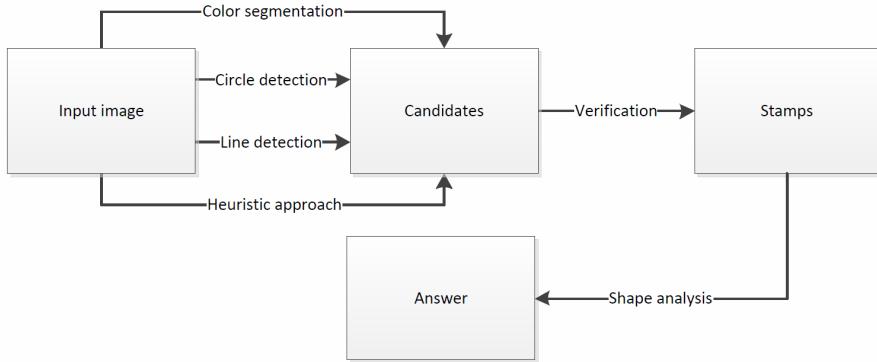


Fig. 1 Workflow diagram

The first step is to detect areas that may contain interesting objects. For this purpose, we look for elementary structures, like lines, circles and any other consistent areas. In case of color images, we perform a color conversion $RGB \rightarrow YC_bC_r$ and work on C_b and C_r planes in an independent manner [2]. Next, we obtain candidate areas which are later verified using a number of features that characterize specific shapes and classify them into stamps and no-stamps (objects with similar shape, yet with different raster features). The further recognition is performed using one of several investigated classifiers.

More detailed information about all stages are presented in the following sections, however in this paper we focus on the low-level image features employed in classification as it seems, according to our opinion, the most challenging problem. Moreover, a few of these steps have already been precisely described in our previous works [2, 3].

2.1 Image Preprocessing

Since RGB space is not optimal in terms of color image segmentation due to the correlation between channels [2], an input image is converted into YC_bC_r representation. It is especially usable because of the red/blue separation. Moreover, it is a native color format for JPEG files. We focus on this color space, since according to our observations, less than 4% of the stamps are represented by colors other than blue or red. According to our previous observations [2] we investigate objects with dimensions not smaller than 5% of shorter edge of input image and not larger than 40% of its longer edge. Both matrices C_b and C_r are subjected to an averaging filtering in order to reduce noise. The filter uses a mask which size is equal to 3% of input image's shorter edge. This ensures favorable results in the aspect of computation cost vs. quality. The next step is a binarization with different thresholds until at least one of the areas meets an assumptions about the size of the stamp. Additionally,

before filling holes inside found area, it is subjected to morphological operations - region growing. Binary image is then labeled and each candidate region is passed to the module responsible for stamp detection. It works on regions extracted from the channel Y on the basis of binary mask being an output from the previous stage.

2.2 Stamp Detection

Different stamp shapes are detected by independent modules (see Fig. 1).

Detection of circles is done using the Circular Hough Transform (CHT) that works on intensity image. At first, the edge detection using Canny detector is performed, then morphological opening in order to eliminate noise is employed. Finally, classical HT is used.

Shapes consisting of straight lines (squares, rectangles, triangles) are extracted using line detector based on Hough Transform. It operates on image containing simple pixel gradients. Firstly, the input image is smoothed in order to remove noise. Then, we select pixels of high intensity (above a certain threshold) and for each one we employ an approach described in [1] which is an alternative to the original Hough approach, where slope-intercept line representation was used.

Any other shapes are detected using heuristic approach presented below. Firstly, the input image with dimensions of $H_{doc} \times W_{doc}$ pixels is converted to grayscale and then binarized using adaptive thresholding, so it contains a number of closed areas. For each such area a contour is extracted and its bounding box is defined. The area of bounding box is given as $A_{box} = H_{box}W_{box}$, where W_{box} and H_{box} are its width and height, respectively. Additionally, we calculate the number of pixels in each area A_p . Candidates that satisfy the following conditions are passed to the further processing: $400 < A_p < 5000$ and $\min\{H_{box}, W_{box}\} > 0.05 * \min\{H_{doc}, W_{doc}\}$ and $\max\{H_{box}, W_{box}\} < 0.40 * \max\{H_{doc}, W_{doc}\}$.

2.3 Stamp/No-Stamp Verification

The verification is performed using a set of 11 object's characteristics captured in the spatial domain. There are 7 direct features: average pixel intensity, intensity standard deviation, median value of intensity, pixel contrast, brightness to contrast ratio, intensity of edges, brightness to edges intensity ratio, and 4 features calculated from coocurrence matrix: contrast, correlation, energy and homogeneity. In order to select a classifier we performed several experiments involving Naive Bayes Classifier (NBC), Linear Logistics Regression (LLR), Multilayer Perceptron (ML) and K-Nearest Neighbors (kNN) with different values of k . The rates of True Positive and False Positive are given in Tab. 1. Based on these results at the stage of stamp/no-stamp verification we selected kNN with $k = 1$ as a method with the lowest FP rate.

Table 1 Classifiers efficiency comparison

Classifier	TP Rate	FP Rate	Classifier	TP Rate	FP Rate
NB	91.8%	49.6%	kNN, k = 1	83.4%	15.4%
LLR	84.1%	37.7%	kNN, k = 2	91.7%	31.4%
MLP	86.5%	26.9%	kNN, k = 3	84.9%	21.3%

2.4 Stamp Shape Classification

Shape analysis module calculates several measures that are later evaluated at the decision stage. It uses the following measures as the elementary features (M_1 - Minor Axis Length, M_2 - Major Axis Length, A_p - Object Area defined by the number of pixels, P - Object Perimeter), which are later employed to build more complex characteristics. These characteristics were chosen intentionally, since they give a maximal discriminative power in terms of shape classification. All of them are normalized to the interval $[0, 1]$.

Roundness is evaluated according to the average value of the three following measures R_1, R_2 and R_3 :

$$R_1 = \left| 1 - \frac{|M_2 - M_1|}{\max(M_1, M_2)} \right|, \quad (1)$$

$$R_2 = \left| 1 - \frac{\left| \frac{\pi \left(\frac{M_1 + M_2}{2} \right)^2}{2} - A_p \right|}{A_p} \right|, \quad (2)$$

$$R_3 = \left| 1 - \frac{|\pi \max(M_1, M_2) - P|}{P} \right|. \quad (3)$$

Squareness is calculated according to the following:

$$S = \left(\frac{\min(\sqrt{A_p}, \frac{P}{4})}{\max(\sqrt{A_p}, \frac{P}{4})} \right)^2. \quad (4)$$

Vertices number is calculated as a function of object's extreme points, i.e. if an object contains three extremes then it is considered to be a triangle. However, binary objects that are extracted from real documents are often noised or distorted hence finding extremes is very difficult or the number of detected extremes can be different from the truth. In order to assess the extremes an input object is binarized, then we select all extreme points (their coordinates) and create a new binary matrix containing zeros. Next, we put ones in the cells that match found coordinates. Further we cluster the points to their groups based on their spatial location. At this stage we

perform a dilation procedure in order to fill gaps and create groups of neighboring pixels. Then, basing of the number of groups, the extremes are detected.

Aspect ratio is calculated as the proportion of object's width and height (in the similar manner to the one presented in Sec. 2.2):

$$R = \frac{M_1}{M_2}. \quad (5)$$

Extent is the ratio of the number of pixels lying in a rectangle that is circumscribed around an object to the total number of pixels belonging to the object itself:

$$E_X = \frac{H_{box}W_{box}}{A_p}. \quad (6)$$

Another measures, that are later employed include ellipticity, elliptic variance, circular variance, triangularity coefficient (based on central moments) and minimal bounding figure - a coefficient defining a ratio between the area of a smallest shape that is wrapped around the object and this object's area. They involve the use of moment invariants μ_{pq} (of second order), where each contour point $i \in 1, \dots, N$ is described by its coordinates $p_i = (x, y)$. The centroid is calculated as a mean of all points coordinates: $\mu = \frac{1}{N} \sum_i p_i$ and the mean radius is equal to $\mu_r = \frac{1}{N} \sum_i \|p_i - \mu\|$.

Ellipticity [11] is calculated using the following formula:

$$I_1 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{00}^4}. \quad (7)$$

The moment for a unit radius circle allows to compute $I_1 = \frac{1}{16\pi^2}$ for a perfect figure. Ellipticity value ranges over $[0, 1]$ and is computed as follows [11]:

$$E_I = \begin{cases} \frac{16\pi^2 I_1}{16\pi^2 I_1} & \text{if } I_1 \leq \frac{1}{16\pi^2} \\ \frac{1}{16\pi^2 I_1} & \text{otherwise} \end{cases}. \quad (8)$$

Circular variance described in [10] is calculated as follows:

$$C_{var} = \frac{1}{N\mu_r^2} \sum_i (\|p_i - \mu\| - \mu_r)^2. \quad (9)$$

Elliptic variance describing ellipses employs the following formula [10]:

$$E_{var} = \frac{1}{N\mu_{rc}} \sum_i (\sqrt{(p_i - \mu)^\top C^{-1}(p_i - \mu)} - \mu_{rc})^2 \quad (10)$$

where: $C = \frac{1}{N} \sum_i (p_i - \mu)(p_i - \mu)^\top$ is a covariance matrix, while $\mu_{rc} = \frac{1}{N} \sum_i \sqrt{(p_i - \mu)^\top C^{-1}(p_i - \mu)}$.

Triangularity also employs moment invariants [11] and for an ideal right-angled triangle is equal to $I_1 = \frac{1}{108}$, thus the triangularity of an analysed object is computed as follows (again, giving values from the interval $[0, 1]$):

$$T_I = \begin{cases} \frac{108I_1}{108I_1} & \text{if } I_1 \leq \frac{1}{16\pi^2} \\ \frac{1}{108I_1} & \text{otherwise} \end{cases}. \quad (11)$$

Minimal Bounding Figure Coefficient - is calculated as a ratio of object's actual area to the area of an ideal figure that is circumscribed on it. It is called minimal bounding figure and the method of calculation is similar to [11].

In the shape analysis we employ also a coefficient A_p related to the object's area understood as the number of pixels belonging to it. The following group of geometrical features is calculated on the basis of object's area, its shortest radius (span) R_s and its longest radius (span) R_l . If the difference $S_{cir} = R_s - R_l$ is small enough (in our case, lower than 10 pixels) we assume that the object has circular shape. On the other hand, an object is considered to be a square if a value of S_{sqr} from the following equation is close to one:

$$S_{sqr} = \frac{A_p}{4R_s^2}. \quad (12)$$

An ellipse is described as a coefficient:

$$S_{eli} = \frac{A_p}{\pi R_s R_l}. \quad (13)$$

If it is close to one, then the object is recognized as an ellipse. In the same manner, the coefficients related to other shapes, like rectangle, triangle and rhombus (which, for classification purposes, is treated as a square) can be calculated respectively, as:

$$S_{rct} = \frac{A_p}{4R_s \sqrt{R_l^2 - R_s^2}}, \quad (14)$$

$$S_{tri} = \frac{A_p \sqrt{3}}{(R_l + R_s)^2}, \quad (15)$$

$$S_{rho} = \frac{A_p \sqrt{R_l^2 - R_s^2}}{2R_l^2 R_s}. \quad (16)$$

At the decision stage a voting classifier uses weights equal to one for all features, except the following, which have weights equal to 3: R , S_{sqr} , S_{rec} , S_{eli} , T_I , Minimal Bounding Triangle, Mininal Bounding Circle, Minimal Bounding Square (from the group of Minimal Bounding Figure).

3 Experiments

During the experiments we evaluated the performance of both stages of processing, namely the stamp detector and stamp's shape classifier. The experiments were performed on our own benchmark database consisting of a number of images collected from the Internet. The details of this database are as follows: documents with no stamps: 294 (41%), with single stamp: 309 (43%), with multiple stamps: 116 (16%) and with logotypes: 367 (51%). Table 2 shows the confusion matrix obtained during experiments performed on 2925 graphical objects extracted from above described documents (a part of them was obtained by geometrical transformations and noising of original ones). There are 1589 stamps and 1336 no-stamps (logos, ornaments, large letters) in the set. They were divided into learning and testing subsets according to the 10-fold cross validation. During experiments the algorithm was able to recognize 2456 objects, while the rest 469 were misclassified. Table 3 shows the confusion matrix for shape classification.

Table 2 Confusion matrix with true positive and false positive indicators

	Stamps	No-Stamp	False Positives	True Positives
Stamps	1326	263	15.4%	83.4%
No-Stamp	206	1130	16.6%	84.6%

Table 3 Confusion matrix for shape analysis module. Total efficiency - 79%.

Analysed shape	Predicted				
	Circle	Ellipse	Square	Triangle	Rectangle
Circle	426	4	14	2	0
Ellipse	104	266	5	5	37
Square	84	6	226	20	66
Triangle	10	12	0	279	20
Rectangle	1	11	0	15	488

The other experiment was devoted to overall performance of the proposed approach. It included the stage of shape detection and classification. It was performed on a set of 719 scanned or photographed documents. For such experiment we obtained the results presented in Tab. 4.

Provided tables prove that the proposed set of algorithms is able to detect and classify most of stamps. However, it can also be noticed that a large part of them is ignored. It comes from the fact that the database consists of many "difficult" cases - in many cases stamps are hardly visible, only partially imprinted and overlap objects such as pictures, post stamps or printed text (see examples in Fig. 2).

Table 4 The performance of the whole algorithm

	Detected	Missed	Accepted (TP)	Rejected (FN)
Stamp	417 / 70.44%	175 / 29.56%	313 / 75.06%	104 / 24.94%
Non-Stamp	377	—	129 / 34.22%	248 / 65.78%

**Fig. 2** Sample stamps correctly recognized and misclassified at the shape analysis stage

There is also a problem of detecting stamps containing small tables or just texts. In such cases it is almost impossible to distinguish them from actual tables or parts of text. The decrease in classification performance is caused by also the fact, that the database contains documents without stamps, documents with objects similar to stamps, or actual stamps that do not meet conditions specified in Sec. 2.2. As it can be seen from the observations, a large part of the documents has a low contrast. Many wrongly detected areas (later rejected by the classifier) are simply the result of strong compression of the input image.

4 Summary

A novel approach for detecting and classifying stamp instances in scanned documents was presented. It incorporates known methods of image processing, pattern recognition as well as some heuristic. The algorithm is multi-stage, consisting of object detection, verification and classification. Selected features are invariant to rotation, scale and translation, hence this approach is independent on the document size and orientation. The experiments performed on a large set of real documents show the acceptable performance of the method. However, the selection of additional features of stamps could give a significant increase in the overall robustness. The potential area of application includes automatic scanning software for paper documents, content-based document retrieval and various postal services.

Acknowledgements. A great number of documents employed in our experiments was provided by Dr. Barbora Micenkova (Dep. of Computer Science, Aarhus University).

References

1. Duda, R.O., Hart, P.E.: Use of the Hough Transformation to Detect Lines and Curves in Pictures. *Communications of the ACM* 15(1) (1972)
2. Forczmański, P., Frejlichowski, D.: Robust Stamps Detection and Classification by Means of General Shape Analysis. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part I. LNCS, vol. 6374, pp. 360–367. Springer, Heidelberg (2010)
3. Forczmański, P., Frejlichowski, D.: Efficient stamps classification by means of point distance histogram and discrete cosine transform. In: Burduk, R., Kurzyński, M., Woźniak, M., Żołnierk, A. (eds.) Computer Recognition Systems 4. AISC, vol. 95, pp. 327–336. Springer, Heidelberg (2011)
4. Frejlichowski, D.: An Experimental Comparison of Seven Shape Descriptors in the General Shape Analysis Problem. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010. LNCS, vol. 6111, pp. 294–305. Springer, Heidelberg (2010)
5. Hassanzadeh, S., Pourghassem, H.: A Novel Logo Detection and Recognition Framework for Separated Part Logos in Document Images. *Australian Journal of Basic and Applied Sciences* 5(9), 936–946 (2011)
6. Jain, R., Doermann, D.: Logo Retrieval in Document Images. In: 2012 10th IAPR International Workshop on Document Analysis Systems (DAS), pp. 135–139 (2012)
7. Kleban, J., Xie, X., Ma, W.Y.: Spatial Pyramid Mining for Logo Detection in Natural Scenes. In: Proc. of the International Conf. on Multimedia and Expo., Hannover (2008)
8. Micenková, B., van Beusekom, J.: Stamp Detection in Color Document Images. In: International Conf. on Document Analysis and Recognition (ICDAR), pp. 1125–1129 (2011)
9. Okarma, K., Mazurek, P.: Application of shape analysis techniques for the classification of vehicles. In: Mikulski, J. (ed.) TST 2010. CCIS, vol. 104, pp. 218–225. Springer, Heidelberg (2010)
10. Peura, M., Iivarinen, J.: Efficiency of Simple Shape Descriptors. *Aspects of Visual Form*, 443–451 (1997)
11. Rosin, P.L.: Measuring shape: ellipticity, rectangularity, and triangularity. *Machine Vision and Applications* 14(3), 172–184 (2003)
12. Roy, P.P., Pal, U., Lladós, J.: Document seal detection using GHT and character proximity graphs. *Pattern Recognition* 44(6), 1282–1295 (2011)
13. Seiden, S., Dillencourt, M., Irani, S., Borrey, R., Murphy, T.: Logo Detection in Document Images. In: Proc. of the International Conference on Imaging Science, Systems, and Technology, pp. 446–449 (1997)

Stochastic Approximation to Reconstruction of Vector-Valued Images

Dariusz Borkowski

Abstract. In this paper we present a new method of reconstruction of vector-valued images with additive Gaussian noise. In order to solve this inverse problem we use stochastic differential equations with reflecting boundary. The reconstruction algorithm is based on Euler's approximations of solutions of such equations. We consider Euler scheme with random terminal time and controlled parameter of diffusion which is driven by geometry of \mathbf{R}^n -valued noisy image. Our numerical experiments show that the new approach gives very good results and compares favourably with deterministic partial differential equation methods.

1 Introduction

Let D be a bounded, convex domain in \mathbf{R}^2 , $u : \overline{D} \rightarrow \mathbf{R}^n$ be an original image and $u_0 : \overline{D} \rightarrow \mathbf{R}^n$ be the observed image of the form $u_0 = u + \eta$, where η stands for a white Gaussian noise (added independently to all coordinates). We assume that u and u_0 are appropriately regular. We are given u_0 , the problem is to reconstruct u . This is a typical example of an inverse problem [1].

Stochastic methods of image reconstruction are generally based on the Markov field theory, however some papers [3, 4, 7, 13, 14] involve advanced tools of stochastic analysis such as stochastic differential equations. The weakest point of this approach in the case of image denoising is the necessity of using Monte Carlo method. In particular, we have to do multiple simulations of trajectories of the diffusion process. Euler's approximation [11] is a classical method of diffusion simulations. This scheme gives good results only for small time-step discretization, but unfortunately reconstruction takes a very long time. In [3] the Euler scheme was improved for applications to image processing by adding a controlled parameter. This new

Dariusz Borkowski

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University,
Chopina 12/18, 87-100 Toruń, Poland

e-mail: dbor@mat.umk.pl

numerical scheme considered for a constant terminal time T is called a modified diffusion (in short MD) and gives good results for long time-step discretization and reconstruction is about 50 times faster than with the Euler's approximation. In [4] the authors introduced a numerical scheme based on Euler's approximations with random terminal time. They considered the modified diffusion method with terminal time which depends on the geometry of the reconstructed image and therefore it is random. These modifications of the classical Euler scheme lead to the algorithm of denoising gray level images: modified diffusion with random terminal time (in short MDRRT), which compares favourably with modified diffusion method [3] and other classical denoising PDE filters.

A novel look in [4] on the reconstruction problem with the use of stochastic Euler approximation was fruitful and gave encouraging results for gray level images. The idea of this paper is to generalize these results to images with values in \mathbf{R}^n , in particular to colour images.

2 Mathematical Preliminaries

The reconstruction of images is based on two advanced tools of stochastic analysis: stochastic differential equations (in order to model image diffusion) and Skorokhod problem (in order to constrain the diffusion to image domain).

First we will define the Skorokhod problem. Let $D \subset \mathbf{R}^n$ be a domain with closure \overline{D} and boundary ∂D . Let $T > 0$ and by $\mathbf{C}([0, T]; \mathbf{R}^n)$ we denote a set of continuous functions $f : [0, T] \rightarrow \mathbf{R}^n$.

Definition 1. Let $y \in \mathbf{C}([0, T]; \mathbf{R}^n)$, $y_0 \in \overline{D}$. A pair $(x, k) \in \mathbf{C}([0, T]; \mathbf{R}^{2n})$ is said to be a solution to the Skorokhod problem associated with y and D if

1. $x_t = y_t + k_t$, $t \in [0, T]$,
2. $x_t \in \overline{D}$, $t \in [0, T]$,
3. k is a function with bounded variation $|k|$ on $[0, T]$, $k_0 = 0$ and

$$k_t = \int_0^t n_s d|k|_s, |k|_t = \int_0^t 1_{\{x_s \in \partial D\}} d|k|_s, \quad t \in [0, T],$$

where $n_s = n(x_s)$ is an inward normal unit vector at $x_s \in \partial D$.

It is known that if D is a convex set, then there exists a unique solution to the Skorokhod problem [12].

Definition 2. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space.

1. An n -dimensional stochastic process $X = \{X_t; t \in [0, T]\}$ is a parametrised collection of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with values in \mathbf{R}^n .

For each fixed $\omega \in \Omega$ the function $X_t(\omega)$, $t \in [0, T]$ is called a trajectory of X and is denoted by $X(\omega)$.

2. A filtration $(\mathcal{F}_t) = \{\mathcal{F}_t; t \in [0, T]\}$ is a nondecreasing family of sub- σ -fields of \mathcal{F} , i.e. $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $0 \leq s < t \leq T$.
By (\mathcal{F}_t^X) we denote a filtration generated by process X , i.e.
 $\mathcal{F}_t^X = \sigma(X_s; 0 \leq s \leq t)$.
3. A stochastic process X is adapted to the filtration (\mathcal{F}_t) (X is (\mathcal{F}_t) adapted) if for each $t \in [0, T]$, X_t is a \mathcal{F}_t - measurable random variable.

Definition 3. Let Y be (\mathcal{F}_t) adapted process with continuous trajectories, $Y_0 \in \overline{D}$. We say that a pair (X, K) of (\mathcal{F}_t) adapted processes is a solution to the Skorokhod problem associated with Y and D , if for almost every $\omega \in \Omega$, $(X(\omega), K(\omega))$ is a solution to the Skorokhod problem associated with $Y(\omega)$ and D .

In what follows, by $W = \{W_t; t \in [0, T]\}$ we shall denote a Wiener process starting from zero. We assume that we are given a point $x_0 \in \overline{D}$ and some function $\sigma : \mathbf{R}^n \rightarrow \mathbf{R}^n \times \mathbf{R}^m$.

Definition 4. Let Y be an (\mathcal{F}_t) adapted process. A pair $(X, K^{\overline{D}})$ of (\mathcal{F}_t) adapted processes is called a solution to reflected SDE

$$X_t = x_0 + \int_0^t \sigma(X_s) dW_s + K_t^{\overline{D}}, \quad t \in [0, T], \quad (1)$$

if $(X, K^{\overline{D}})$ is a solution to the Skorokhod problem associated with

$$Y_t = x_0 + \int_0^t \sigma(X_s) dW_s, \quad t \in [0, T] \text{ and } D.$$

The process X is called the process with reflection. The proof of existence and uniqueness of the solution to reflected SDEs can be found in [12].

3 Reconstruction of Gray Level Images

We suppose for a while that the image is given by a function defined on the whole plane. Put $X_t = W_t^x$, $t \in [0, T]$, where W^x is a two-dimensional Wiener process starting from $x \in \overline{D}$. Then

$$\mathbf{E}[u_0(X_T)] = \int_{\mathbf{R}^2} \frac{1}{2\pi T} e^{-\frac{|x-y|^2}{2T}} u_0(y) dy = \int_{\mathbf{R}^2} G_{\sqrt{T}}(x-y) u_0(y) dy, \quad (2)$$

where $G_{\sqrt{T}}(x) = \frac{1}{2\pi T} e^{-\frac{|x|^2}{2T}}$ is a two-dimensional Gaussian mask.

The reconstructed pixel $u(x)$ is defined as the mean value $\mathbf{E}[u_0(X_T)]$. Therefore, by (2) the image is the convolution of the noise image with the two-dimensional Gaussian mask.

Since we want to consider the image as a function defined on the bounded convex set, we have to introduce a new assumption on the process X . It is natural to assume that the process X is a stochastic process with reflection with values in \overline{D} . In this case the process X is given by a Wiener process with reflection, i.e. it can be written as $X_t = W_t^x + K_t^{\overline{D}}$ (see Definition 4).

The above model removes noise and blurs edges. Following [8, 16] we provide a construction of an anisotropic diffusion model, where noise is removed and image has sharp edges. These conditions may be achieved by imposing

$$X_t = x + \int_0^t \begin{bmatrix} -\frac{(G_\gamma * u_0)_{x_2}(X_s)}{|\nabla(G_\gamma * u_0)(X_s)|}, 0 \\ \frac{(G_\gamma * u_0)_{x_1}(X_s)}{|\nabla(G_\gamma * u_0)(X_s)|}, 0 \end{bmatrix} dW_s + K_t^{\overline{D}}, \quad (3)$$

where $u_{x_i}(y) = \frac{\partial u}{\partial x_i}(y)$ and

$$u(x) = \mathbf{E}[u_0(X_T)].$$

To avoid false detections due to noise, u_0 is convolved with a Gaussian kernel $G_\gamma(x) = \frac{1}{2\pi\gamma^2} e^{-\frac{|x|^2}{2\gamma^2}}$ (in practice a 3×3 Gaussian mask).

3.1 Euler's Approximation

Consider the following numerical scheme

$$X_0^m = X_0, \quad X_{t_k}^m = \Pi_{\overline{D}}[X_{t_{k-1}}^m + \sigma(X_{t_{k-1}}^m)(W_{t_k} - W_{t_{k-1}})], \quad k = 1, 2, \dots, m, \quad (4)$$

where $t_k = kh$, $h = \frac{T}{m}$, $k = 0, 1, \dots, m$ and $\Pi_{\overline{D}}(x)$ denotes a projection of x on the set \overline{D} . Since D is convex, the projection is unique.

Theorem 1. Let $(X, K^{\overline{D}})$ be the solution to the reflected SDE (1). If there exists $C > 0$ such that $\|\sigma(x) - \sigma(y)\|^2 \leq C|x - y|^2$, then

$$\lim_{m \rightarrow +\infty} |X_T^m - X_T| = 0 \text{ almost surely.}$$

The proof of the above theorem can be found in [11].

3.2 Modified Diffusion

The numerical scheme (4) gives good results, but only with a small value of the time-step parameter $h = \frac{T}{m}$ (for example $h = 0.05$). Calculating the mean value using Monte Carlo method for small h is not effective and takes a long time. To omit this problem, we improve the scheme (4) by adding a controlled parameter p [3].

$$X_0^m = X_0, \quad H_{t_k}^m = \Pi_{\overline{D}}[X_{t_{k-1}}^m + \sigma(X_{t_{k-1}}^m)(W_{t_k} - W_{t_{k-1}})],$$

$$X_{t_k}^m = \begin{cases} H_{t_k}^m, & \text{if } \Theta, \\ X_{t_{k-1}}^m, & \text{elsewhere,} \end{cases} \quad k = 1, 2, \dots, m, \quad (5)$$

where by Θ we mean the condition $|(G_\gamma * u_0)(H_{t_k}^m) - (G_\gamma * u_0)(X_{t_{k-1}}^m)| < p$.

Note that the parameter $p > 0$ guarantees that if the image exhibits a strong gradient then the process X^m diffuses as a process with small value of the parameter h and at locations where variations of the brightness are small, the process X^m can diffuse with a large value of h (for example $h = 4$).

For small h or $p = +\infty$ (in practice $p > 255$) the numerical scheme (5) is equivalent to the scheme (4).

3.3 Modified Diffusion with Random Terminal Time

At locations where gradient is large in all directions it is possible that condition Θ does not hold as many times as we would expect. To avoid this we propose the following modification [4]:

$$X_0^m = X_0, \quad H_{t_k}^m = \Pi_{\overline{D}}[X_{t_{k-1}}^m + \sigma(X_{t_{k-1}}^m)(W_{t_k} - W_{t_{k-1}})],$$

$$X_{t_k}^m = \begin{cases} H_{t_k}^m, & \text{if } \Theta, \\ X_{t_{k-1}}^m, & \text{elsewhere,} \end{cases} \quad k = 1, 2, \dots, \tau_m, \quad (6)$$

where $\tau_m = \min\{k; k \geq m \text{ and } \Theta \text{ is true } m \text{ times}\}$.

Terminal time τ_m guarantees that the numerical simulation of the diffusion trajectory gives at least m values of $X_{t_k}^m$ which differ from the value in the previous step. Observe that the scheme (6) works well only if the model of the digital image $G_\gamma * u_0$ is continuous. In practice, we can use a linear interpolation to get the value of the image $G_\gamma * u_0$, for any point $x \in \overline{D}$.

4 Reconstruction of Vector-Valued Images

Now we concentrate on images with values in \mathbf{R}^3 . A very common idea to restore vector-valued images is to use scalar diffusion on each channel of a noisy image. But one quickly notices that this scheme is useless, since each image channel evolves independently with different smoothing geometries. To avoid this blending effect, the regularization process has to be driven in a common and coherent way for all the vector image channels. In order to execute that we use Di Zenzo geometry [5, 6].

Let $u : D \rightarrow \mathbf{R}^3$ be a vector-valued image and $x \in D$ be fixed. Consider the function $F_x : V \rightarrow \mathbf{R}$, $F_x(v) = \left| \frac{\partial u}{\partial v}(x) \right|^2$, where $V = \{v \in \mathbf{R}^2; |v| = 1\}$. We are interested in finding the arguments $\theta_+(u, x), \theta_-(u, x)$ and corresponding values $\lambda_+(u, x) = F_x(\theta_+(u, x)), \lambda_-(u, x) = F_x(\theta_-(u, x))$ which maximize and minimize the function F_x , respectively.

Note that F_x can be rewritten as $F_x(v) = F_x([v_1, v_2]^T) = v^T \mathbf{G}(x)v$, where

$$\mathbf{G}(x) = \begin{bmatrix} \sum_{i=1}^3 \left(\frac{\partial u_i}{\partial x_1}(x) \right)^2, & \sum_{i=1}^3 \frac{\partial u_i}{\partial x_1}(x) \frac{\partial u_i}{\partial x_2}(x) \\ \sum_{i=1}^3 \frac{\partial u_i}{\partial x_1}(x) \frac{\partial u_i}{\partial x_2}(x), & \sum_{i=1}^3 \left(\frac{\partial u_i}{\partial x_2}(x) \right)^2 \end{bmatrix}.$$

The interesting point about $\mathbf{G}(x)$ is that its positive eigenvalues $\lambda_+(u, x), \lambda_-(u, x)$ are the maximum and the minimum of F_x while the orthogonal eigenvectors $\theta_+(u, x)$ and $\theta_-(u, x)$ are the corresponding variation orientations.

Three different choices of vector gradient norms $N(u, x)$ have been proposed in the literature $N_1(u, x) = \sqrt{\lambda_+(u, x)}, N_2(u, x) = \sqrt{\lambda_+(u, x) - \lambda_-(u, x)}, N_3(u, x) = \sqrt{\lambda_+(u, x) + \lambda_-(u, x)}$. In presented examples we have used $N(u, x) = \sqrt{\lambda_+(u, x)}$ as a natural extension of the scalar gradient norm viewed as the value of maximum variations.

Replacing in equation (3) $|\nabla(u, x)|$ and $[u_{x_1}(x), u_{x_2}(x)]^T$ respectively by $N(x, u)$ and $\theta_+(u, x) = [\theta_+^1(u, x), \theta_+^2(u, x)]^T$ we obtain the following model of anisotropic diffusion for vector-valued images:

$$X_t = x + \int_0^t \begin{bmatrix} -\frac{\theta_+^1(G_\gamma * u_0, X_s)}{N((G_\gamma * u_0)(X_s))}, 0 \\ \frac{\theta_+^2(G_\gamma * u_0, X_s)}{N((G_\gamma * u_0)(X_s))}, 0 \end{bmatrix} dW_s + K_t^D, \quad (7)$$

where

$$u(x) = \mathbf{E}[u_0(X_T)].$$

4.1 Vector-Valued Modified Diffusion with Random Terminal Time

Considering in condition Θ the L^2 norm we have the following numerical scheme for vector-valued images (in short VMDRTT):

$$X_0^m = X_0, \quad H_{t_k}^m = \Pi_{\overline{D}}[X_{t_{k-1}}^m + \sigma(X_{t_{k-1}}^m)(W_{t_k} - W_{t_{k-1}})],$$

$$X_{t_k}^m = \begin{cases} H_{t_k}^m, & \text{if } \Theta, \\ X_{t_{k-1}}^m, & \text{elsewhere,} \end{cases} \quad k = 1, 2, \dots, \tau_m, \quad (8)$$

where

$$\sigma(X_{t_{k-1}}^m) = \begin{bmatrix} -\frac{\theta_+^1(G_\gamma * u_0, X_{t_{k-1}}^m)}{N((G_\gamma * u_0)(X_{t_{k-1}}^m))}, 0 \\ \frac{\theta_+^2(G_\gamma * u_0, X_{t_{k-1}}^m)}{N((G_\gamma * u_0)(X_{t_{k-1}}^m))}, 0 \end{bmatrix},$$

$\tau_m = \min\{k; k \geq m \text{ and } \Theta \text{ is true } m \text{ times}\}$ and by Θ we mean the condition

$$|(G_\gamma * u_0)(H_{t_k}^m) - (G_\gamma * u_0)(X_{t_{k-1}}^m)| < p,$$

where $|(x_1, x_2, x_3)| = \sqrt{x_1^2 + x_2^2 + x_3^2}$.

The picture in Fig. 1 presents a comparison of reconstruction results received with using the classical Euler's approximation [11] with long and short time-step discretization and VMDRTT (8). Figures c) d) e) are results of the reconstruction with the same value of diffusion parameter T . The reconstruction time of VMDRTT

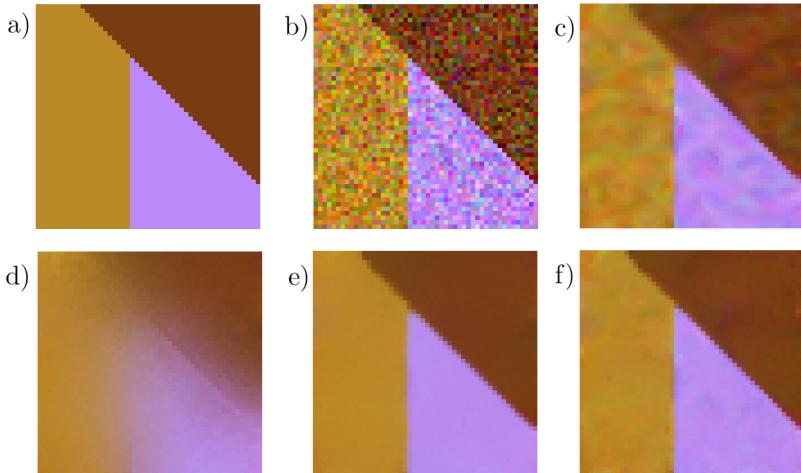


Fig. 1 a) Original image: 50×50 pixels b) Noisy image: standard deviation of the noise $\rho = 30$ c) Euler's scheme: $T = 56$, $h = 0.1$ (540 seconds) d) Euler's scheme: $T = 56$, $h = 4$ (13 seconds) e) VMDRTT: $T = 56$, $h = 4$ (16 seconds), $p = 30$ f) VMDRTT: $T = 16$, $h = 4$, $p = 30$ (5 seconds).

Table 1 SSIM

Standard deviation	$\rho = 20$	$\rho = 30$	$\rho = 40$
PM	0.8630	0.8105	0.7713
TV	0.8792	0.8335	0.7986
VMDRTT	0.8826	0.8438	0.8056

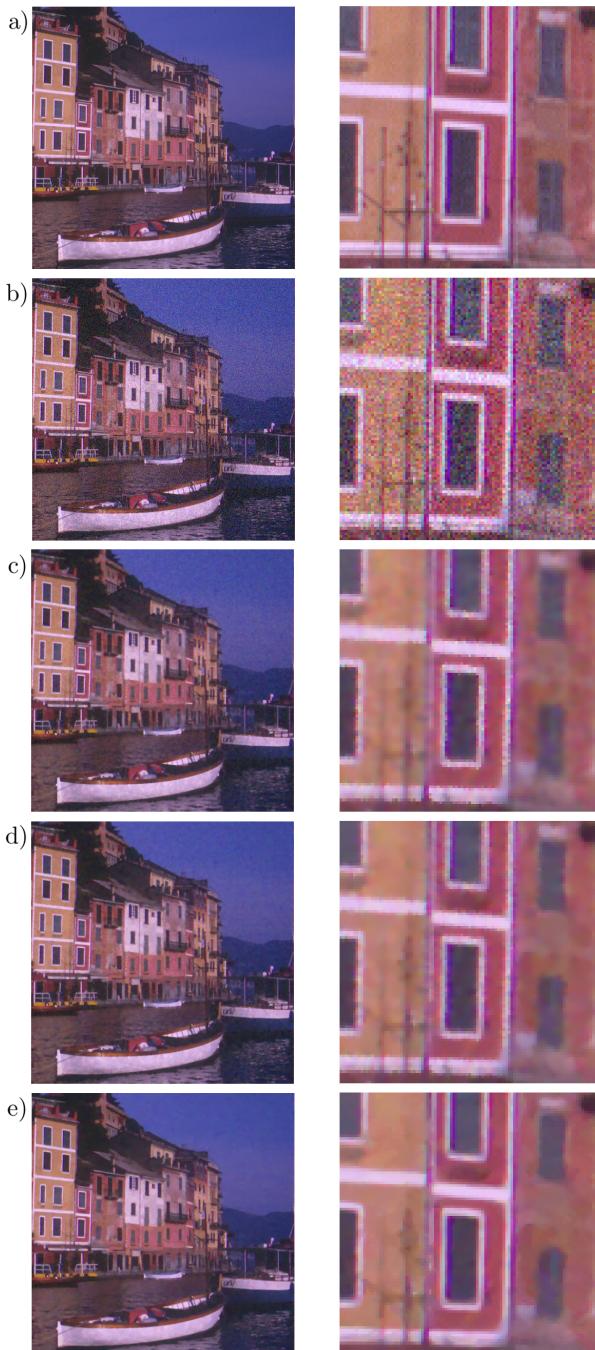


Fig. 2 a) Original image: 512×512 pixels b) Noisy image: $\rho = 20$ c) PM: SSIM=0.8630 d) TV: SSIM=0.8792 e) VMDRTT: SSIM=0.8826. Left: full images; right: a fragment chosen around the two windows surrounded by the darkest red wall.

was substantially reduced and the result is much better. The image c) is comparable to f) which is a result of VMDRTT with short diffusion parameter T . Note that we can obtain comparable results while reducing the time of the reconstruction by two orders – it is about 100 times faster.

5 Experimental Results

Some measures of quality for our evaluation experiments regarding VMDRTT and classic PDE methods: total variation [9] (in short TV) and Perona-Malik [8] (in short PM) for colour images [5, 10] are presented in Table 1 and Fig. 2. The results refer to RGB image *portofino* corrupted (channels independently) with the Gaussian noise with standard deviation ρ . Noisy images have been reconstructed with vector analysis in RGB space. The maximum values of Structural Similarity Index (in short SSIM) are given in the table. Definition of SSIM error in gray scale can be found in [15]. In order to count SSIM in RGB color space we apply SSIM measure to each individual color component and next we average the result [2]. Parameters of SSIM were set to the default values as recommended by [15].

When comparing the figures one can observe that the image created by the stochastic method is visually more pleasant. The reason for this is that PDE methods show clear evidence of a block image, but this stair-case effect is reduced in our algorithm. Moreover, an analysis of the measures of image quality shows that VMDRTT method performs better.

6 Conclusion

In this paper we have presented a new colour denoising method based on Euler's approximation of reflected SDEs. The obtained results demonstrate the efficiency of the proposed approach, compared with the classical Euler scheme.

References

1. Aubert, G., Kornprobst, P.: Mathematical problems in image processing. Springer, New York (2002)
2. Astrom, F., Felsberg, M., Lenz, R.: Color Persistent Anisotropic Diffusion of Images. In: Proceedings of the 17th Scandinavian Conference on Image analysis (SCIA), pp. 262–272 (2011)
3. Borkowski, D.: Modified diffusion to image denoising. Adv. Soft Comp. 45, 92–99 (2007)
4. Borkowski, D.: Euler's Approximations to Image Reconstruction. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 30–37. Springer, Heidelberg (2012)

5. Deriche, R., Tschumperlé, D.: Diffusion PDE's on vector-valued images: local approach and geometric viewpoint. *IEEE Signal Processing Magazine* 19(5), 16–25 (2002)
6. Di Zenzo, S.: A note on the gradient of a multi-image. *Comput. Vis. Graph. Image Process.* 33(1), 116–125 (1986)
7. Juan, O., Keriven, R., Postelnicu, G.: Stochastic Motion and the Level Set Method in Computer Vision: Stochastic Active Contours. *Int. J. Comput. Vision* 69(1), 7–25 (2006)
8. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12(7), 629–639 (1990)
9. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* 60(1-4), 259–268 (1992)
10. Sapiro, G., Ringach, D.L.: Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Trans. Image Process.* 5(11), 1582–1585 (1996)
11. Ślomiński, L.: Euler's approximations of solutions of SDEs with reflecting boundary. *Stoch. Proc. Appl.* 94, 317–337 (2001)
12. Tanaka, H.: Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math. J.* 9(1), 163–177 (1979)
13. Unal, G., Krim, H., Yezzi, A.: Stochastic differential equations and geometric flows. *IEEE Trans. Image Process.* 11(12), 1405–1416 (2002)
14. Unal, G., Ben-Arous, G., Nain, D., Shimkin, N., Tannenbaum, A., Zeitouni, O.: Algorithms for stochastic approximations of curvature flows. In: *Image Processing, Proceedings ICIP 2003*, vol. 2-3, pp. 651–654 (2003)
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 13(4), 600–612 (2004)
16. Weickert, J.: Theoretical Foundations Of Anisotropic Diffusion In Image Processing. *Computing Suppement* 11, 221–236 (1996)

Image Segmentation with Use of Cross-Entropy Clustering*

Marek Śmieja and Jacek Tabor

Abstract. We present an image segmentation approach which is invariant to affine transformation – the result after rescaling the picture remains almost the same as before. Moreover, the algorithm detects automatically the correct number of groups. We show that the method is capable of discovering general shapes as well as small details by the appropriate choice of only two input parameters.

1 Introduction

The image segmentation is a technique which relies on partitioning the image into several regions according to the selected features, for example, the pixel value. It is widely applied in pattern recognition, computer vision and image analysis.

We distinguish various categories of segmentation methods like edge-based or region-based segmentations [4, 5, 11]. One of the most important class includes techniques based on clustering [6, 8]. Such algorithms interpret pixels (or more precisely the groups of pixels) as high dimensional vectors and divide them into groups. Unfortunately, popular clustering methods suffer very often from many drawbacks. k -means algorithm does not detect the correct number of clusters, tends to form groups of similar sizes and is sensitive to affine transformations. More advanced methods as EM (Expectation Maximization) give better results but involve high computational complexity [2, 9, 10].

In this paper a Gaussian cross-entropy clustering (CEC) will be applied to the image segmentation [12]. It can be seen as joining of k -means with EM but is free

Marek Śmieja · Jacek Tabor

Institute of Computer Science, Department of Mathematics and Computer Science,
Jagiellonian University, Lojasiewicza 6, 30-348, Krakow, Poland
e-mail: {marek.smieja, jacek.tabor}@ii.uj.edu.pl

* This research was supported by National Centre of Science (Poland) Grants No. 2011/01/B/ST6/01887).

of all of the aforementioned disadvantages. The algorithm splits data according to the mixture of Gaussian distributions. Its implementation with use of Hartigan approach has similar computational complexity to classical k -means but has two important features – it reduces irrelevant clusters on-line and is invariant to affine transformations.

The illustration of CEC properties and advantages is presented in the next section. Third section contains deeper overview of the method with the practical explanation how to perform segmentation with use of CEC. Our experiments are conducted on gray-scale images selected from Berkeley Segmentation Database [1].

2 Illustration of the Method

In this section we will discuss on the examples the properties of the image segmentation with use of CEC.

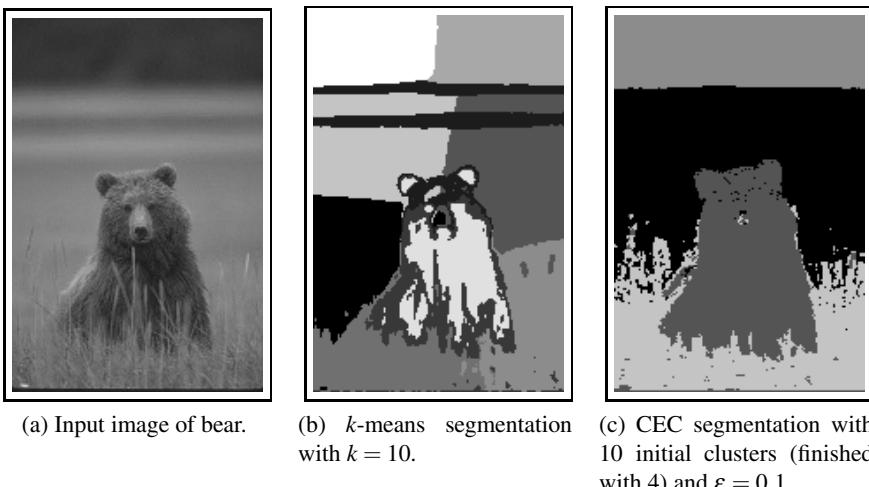


Fig. 1 Detection of the correct number of clusters with use of CEC (c) compared to classical k -means (b)

First experiment concentrates on the demonstration that, in contrast to k -means, CEC is capable of discovering the correct number of groups in the picture.

Example 1. Let us consider an image of a bear showed in Figure 1a. CEC finds the optimal number of segments by the appropriate set of only one input parameter ε – the level when to remove redundant clusters (the meaning of this parameter is explained precisely in the next section). As an input of CEC we also give the initial number of clusters – it is simultaneously changing by the algorithm respectively to the value of ε . We can see in Figure 1c that for ten initial clusters and $\varepsilon = 0.1$ the

algorithm detected four segments which suits well our expectations. The input of k -means algorithm has to include the final number of groups. When we put $k = 10$ then the segmentation results are not satisfactory (see Figure 1b).

The next example presents the comparison of the segmentation with use of CEC and k -means with known parameter k .

Example 2. In the previous experiment the correct number of clusters was not known. Here we assume that this quantity is estimated correctly – for the picture of bear we put $k = 4$. Compare the effect of k -means in Figure 2a with the one obtained with use of CEC in Figure 1c. Let us observe that k -means, contrary to CEC, did not detect properly the image of bear (see Figures 2b and 2c).

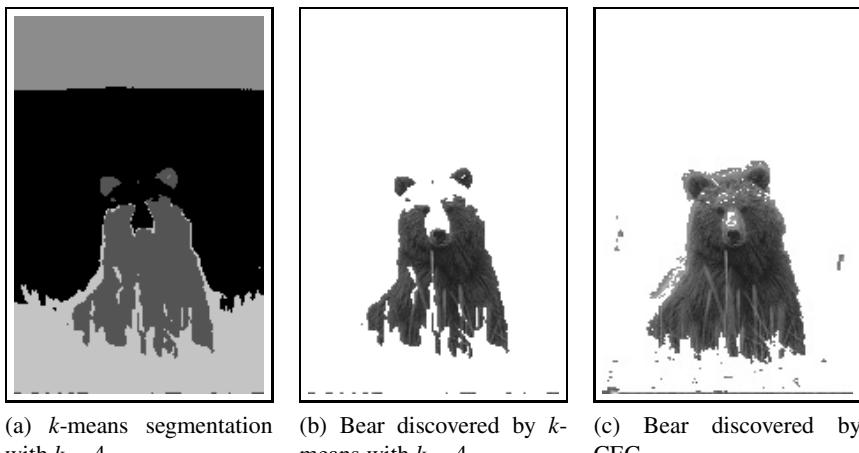


Fig. 2 Segmentation with use of CEC and k -means with $k = 4$

Another important feature of CEC is that it is invariant to any affine transformation – we examine here the segmentation on rescaled image.

Example 3. The clustering methods which rely on metrics are sensitive to scaling. When we resize the image then the segmentation result differs radically from the one with original size. Hence the segmentation based on k -means does not deal well with the scaling.

Let us observe that CEC is free of this drawback since it is based on the data distribution, not on the metric. It is possible to perform various affine transformations – the segmentation result does not differ much from the original one. Clearly, there may appear some differences since the computation on rescaled image takes into account slightly different groups of pixels². However, this is due to the transformation, not the clustering algorithm.

² For a given pixel the input vector is built with use of the adjacent pixel values (pixel window). Therefore, to consider the same groups of pixels on resized picture we should also change the pixel widow.

The effects of segmentations on rescaled image from Figure 3a with use of k -means and CEC are presented in Figure 3b and 3c.

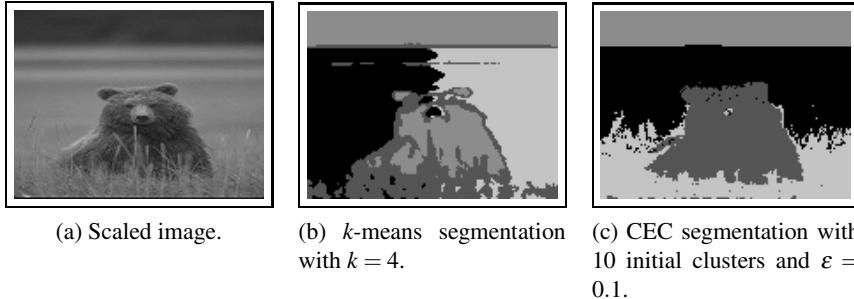


Fig. 3 Segmentation on rescaled image with use of k -means and CEC – compare Figures 3b with 2a and Figure 3c with 1c

3 Performing the Segmentation with Use of CEC

In this section we explain how to perform and control the segmentation with use of CEC. The standard procedure consists of the following steps:

- Select $b \times b$ window around each pixel to use as a feature vector for segmentation (if not stated otherwise then we put $b = 7$).
- Apply PCA (Principal Component Analysis) [7] to these vectors to project them onto their first n principal components (we use $n = 4$). Projection speeds up the clustering algorithm and does not affect much the result of segmentation.
- Supplement every vector with two coordinates of the pixel³. We add the coordinates after performing PCA because color and position represent different quantities and are not comparable values.
- Perform CEC on a sequence of $(n + 2)$ -dimensional vectors.

We explain more precisely the last step: Gaussian CEC clustering [12] relies on finding the pairwise disjoint splitting $(U_i)_i$ of a data-set $U \subset \mathbb{R}^N$ which minimizes the energy function

$$E(U_1, \dots, U_k) = \sum_{i=1}^k p(U_i) \cdot \left[\frac{N}{2} \ln(2\pi\varepsilon) - \ln(p(U_i)) + \frac{1}{2} \ln \det(\Sigma_{U_i}) \right], \quad (1)$$

where $p(U_i) = \frac{\text{card}(U_i)}{\text{card}(U)}$ and Σ_{U_i} denotes the covariance matrix of U_i . The expression (1) corresponds to the mean code length of randomly chosen element with the code optimized for the mixture of k Gaussian distributions.

³ We assume that parts of one segment should not appear in completely different regions of the image.

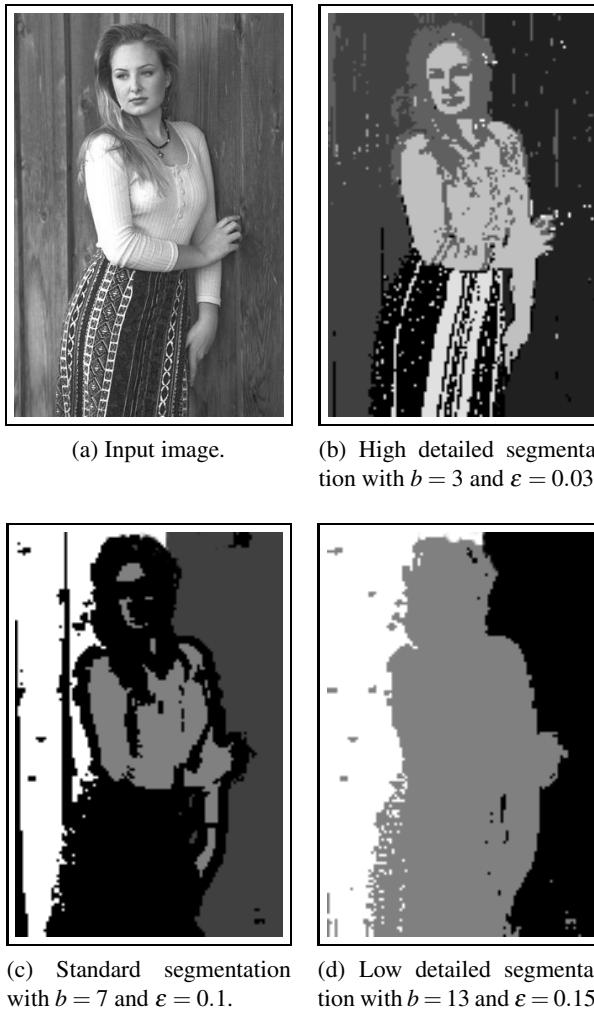


Fig. 4 CEC image segmentation with different values of input parameters

To perform the clustering, we apply the well-known Hartigan approach⁴. Roughly speaking, the procedure relies on iteratively reassigning points to these clusters for which the decrease of energy is maximal. Since CEC introduces the cost of maintaining the cluster (which equals $-p(U_i) \ln(p(U_i))$) the method tends to reduct groups which carry no information. We starts the algorithm with an arbitrary number of initial clusters and remove unnecessary one when its probability is less than a fixed number $\varepsilon \in [0, 1)$. The algorithm can be implemented efficiently to work in comparable time to classical k -means.

⁴ The time of the segmentation process depends strictly on the number algorithm iterations – for an image with the resolution 120×180 one iteration lasts about one second.

The effect of the segmentation depends strictly on the size b of the image window and the parameter ε which describes the level when we decide to remove cluster. The following experiment focuses on the revealing the relationship between the choice of these two parameters and the accuracy of the segmentation. Let us observe that higher values of parameter b makes the image smoother – in consequence we remove details. On the other hand if we are interested in discovering only the most distinctive segments then we should set relatively high value of ε . Let us examine this observation on the following example.

Example 4. We are given an image of the person (see Figure 4a) and we try to discover segments with different precision. We ran the algorithm with 10 initial clusters. The results are presented in Figure 4b, 4c, 4d. We can see that Figure 4b ($\varepsilon = 0.03$ and $b = 3$) reveals all details – hair, eyes, mouth and even the dress pattern while Figure 4d ($\varepsilon = 0.15$ and $b = 13$) finds only the general shape of the person.

4 Conclusion

We presented the application of cross-entropy clustering method in the image segmentation. The main advantage of CEC is that it automatically detects the correct number of clusters. Moreover, it is invariant to affine transformations – the segmentation results after rescaling the image remain almost the same as before scaling. The method with the appropriate choice of only two input parameters is capable of performing the segmentation with various precision. We can conduct general segmentation as well as the detailed one.

References

1. The Berkeley Segmentation Database, <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds>
2. Celeux, G., Govaert, G.: Pattern Recognition 28(5), 781–793 (1995)
3. Cover, T.M., Thomas, J.A., Wiley, J., et al.: Elements of information theory, vol. 6. Wiley Online Library (1991)
4. Deng, Y., Manjunath, B.S.: IEEE Trans. Pattern Anal. Machine Intell. 23(8), 800–810 (2001)
5. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis. John Wiley & Sons, New York (1979)
6. Jain, A.K., Murty, M.N., Flynn, P.J.: ACM Computing Surveys 31(3), 264–323 (1999)
7. Jolliffe, I.: Principal component analysis. Encyclopedia of Statistics in Behavioral Science (2002)
8. Ma, Y., Derksen, H., Hong, W., Wright, J.: IEEE Transactions on Pattern Analysis and Machine Intelligence 29(9), 1546–1562 (2007)

9. McLachlan, G.J., Krishnan, T.: The EM algorithm and extensions, vol. 274. Wiley, New York (1997)
10. McNicholas, P.D., Murphy, T.B.: Statistics and Computing 18(3), 285–296 (2008)
11. Pei, S.C., Ding, J.J.: ICASSP 3, 357–360 (2003)
12. Tabor, J., Spurek, P.: Cross-entropy clustering (2012),
<http://arxiv.org/pdf/1210.5594.pdf>

Detection of Disk-Like Particles in Electron Microscopy Images

P. Spurek, J. Tabor, and E. Zajac*

Abstract. Quantitative and qualitative description of particles is one of the most important tasks in the Electron Microscopy (EM) analysis. In this paper, we present an algorithm for identifying ball-like nanostructures of gahnite in the Transmission Electron Microscopy (TEM) images. Our solution is based on the cross-entropy clustering which allows to count and measure disk-like objects which are not necessary disjoint or with not smooth borders.

1 Introduction

In the recent years, describing the structure of nanomaterials with measure from a few to several hundred nanometers become very important in particular in medicine and biology [3, 4, 9]. Ball-shaped nanoparticles of gahnite with a biologically active layer adsorbed to their surface have the potential to be used as a biological nanosensors [6, 13]. The geometry and the size of a gahnite nanoparticle is very important as it determines its ability to be injected into blood vessel as biological nanosensor. One of the most important and most reliable techniques for identifying the nature and form of nanomaterials is Transmission Electron Microscopy (TEM). In the case of gahnite nanoballs, the quality of nanomaterial produced depend on the number and size of ball-like nanoparticles.

P. Spurek · E. Zajac

Institute of Mathematics, Jan Kochanowski University, Świętokrzyska 15,
25-406 Kielce, Poland

e-mail: {przemyslaw.spurek, ezajac}@ujk.edu.pl

J. Tabor

Faculty of Mathematics and Computer Science, Jagiellonian University,
Łojasiewicza 6, 30-348 Kraków
e-mail: jacek.tabor@im.uj.edu

* The work of this author was supported by the European Regional Development Fund within the Innovative Economy Operational Programme 2007-2013 (UDA-POIG.01.03.01-14-071/08-07).

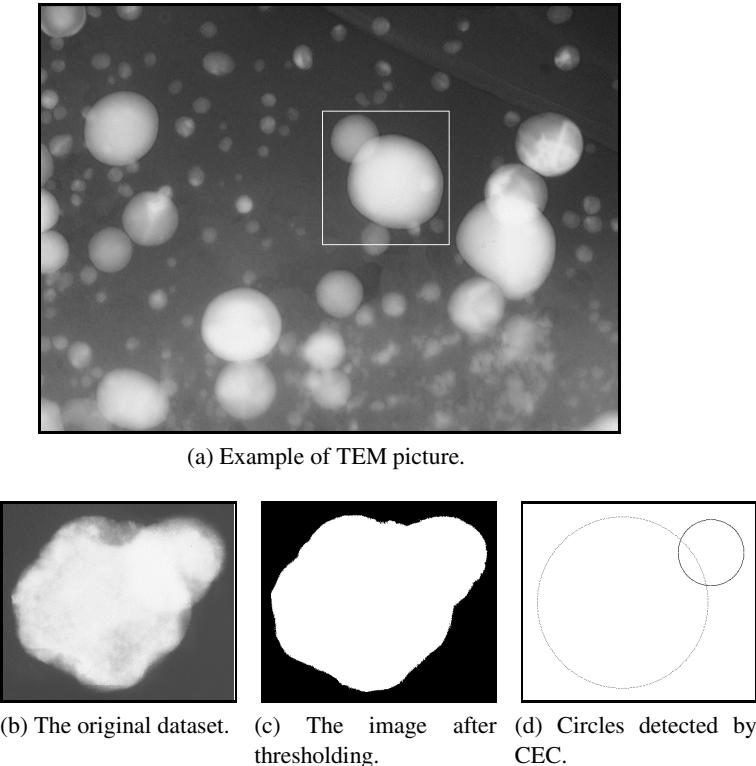


Fig. 1 The results of the CEC algorithm

In this paper we present a method for estimating the number and size of the gahne nanoballs that are observed in TEM images. The nanomaterial analyzed consists of nanocrystalline structures and amorphous substance. The amorphous substance, as shown in Fig. 1, and not completely regular shapes of "nanoball" (2) imply that the standard image processing techniques, which use the Hough Transform (HT) [1, 5, 7, 14], are not sufficient. The basic idea of HT circle detection is based on analyzing edges in images. However, in our case objects in the picture do not have smooth borders and in many cases the "circles" are not disjoint (Fig. 1a) which causes difficulties in determining edges.

On the other hand, due to various types of image deformation caused by specific construction of tested materials and measurement inaccuracies, nanocrystals (similar to 3D balls) sometimes do not look like disks in images. In such cases, the circle is deformed and looks like an ellipse. Moreover, we have to deal with the noise that makes it difficult to analyze images. Fig. 1a presents examples of the TEM pictures.

This paper presents a method based on the cross-entropy clustering [11] (CEC) instead of HT. This method has a few advantages over the standard approach. First of all, we do not need to extract edges. Instead of this we apply simple

thresholding based on the Otsu algorithm [8]. Moreover, we can find elements which are not exactly a disc. The results of our investigation can be applied in the case of nanomaterials with nanoparticles with approximately spherical shapes.

Let us discuss the contents of this paper. In the first part of our work we briefly describe the CEC algorithm and present its advantages and disadvantages in the context of disks detection (in the case of TEM images). In the third section we present practical implementation of our approach.

2 CEC Approach to Circle Detection

In this section we give a short introduction to Spherical CEC² in the case of disks detection, for more detailed explanation we refer the reader to [11]. Let $Y = \{y_1, \dots, y_n\}$ be arbitrary given dataset³. Our goal is to split the data set into k disjoint subsets Y_1, \dots, Y_k (the number of groups is unknown), such that each cluster is well approximated by a disc. Spherical cross-entropy clustering divides the dataset into groups by trying to minimize the "energy function" given by:

$$E((Y_i)_{i=1}^k) = \sum_{i=1}^k p(Y_i) \cdot [\frac{N}{2} \ln(2\pi e/N) - \ln(p(Y_i)) + \frac{N}{2} \ln(\text{Tr}(\Sigma_{Y_i}))], \quad (1)$$

where $p(Y_i) = \text{card}(Y_i)/\text{card}(Y)$ denotes the probability of choosing Y_i , Σ_{Y_i} is covariance matrix of Y_i and m_{Y_i} denotes the mean of Y_i .

Since CEC is a generalization of the classical k-means method we use Hartigans approach to minimize equation (1). Full description of the algorithm can be found in [11]. The most powerful properties of this method is that it simultaneously deletes unnecessary clusters and is scale invariant, that is it well detects disks of various sizes.

In the case of a set containing k disjoint circles (in \mathbb{R}^2), we obtain k clusters containing them where each of group is described by the mean and covariance matrix. Making use of [11], one can deduce the following properties of spherical CEC:

- Let a set $Y \subset \mathbb{R}^2$ be the sum of $k \in \mathbb{N}$ disjoint disks Y_1, \dots, Y_k . Then for any other division $\bar{Y}_1, \dots, \bar{Y}_k$ of Y we have:

$$E((\bar{Y}_i)_{i=1}^k) > E((Y_i)_{i=1}^k).$$

- Let $Y \subset \mathbb{R}^2$ be a disc-like dataset. Then the optimal disc describing Y is approximately given by

$$Y \approx \mathcal{B}(\text{m}_Y, \sqrt{2\text{Tr}(\Sigma_Y)}).$$

As was mentioned earlier, in the case of a TEM picture we need to deal with shapes which are not exactly disks (see Fig. 1(b)). The second problem with the TEM

² Since we use only the Spherical version of CEC, we will use abbreviation CEC to denote Spherical CEC.

³ In our case the set of coordinates of pixels which are inside circles in the picture.

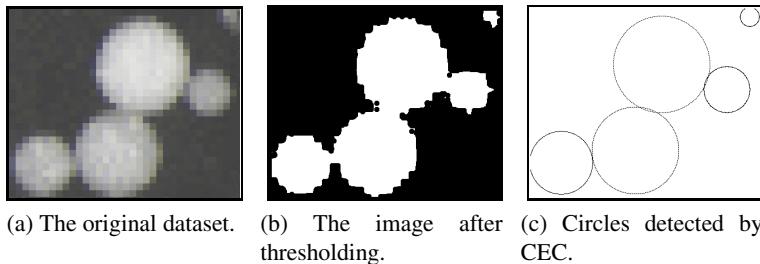


Fig. 2 The results of the CEC algorithm

pictures is that edges can be not smooth (Fig. 4a). In fact, this is another reason why we can not easily use the HT method to analyze this kind of images. Instead of edge detection, CEC use thresholding which is simpler and needs fewer parameters. Fig. 2(c) presents the results of the CEC algorithm in such a case.

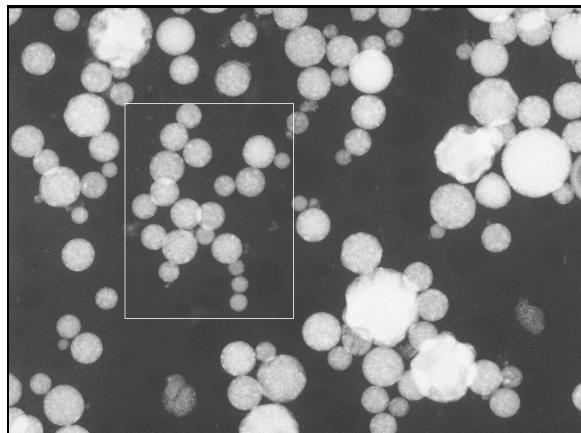
As we see, in our case the shape of edges is not as important as in the classical methods. Since CEC is resistant to low distortion of the set, we can even radically shrink the image before the algorithm is started (Fig. 2a). Even in this situation, CEC detects clusters correctly. A smaller version of the image does not contain all pixels from the original one, so small circles can disappear. Nevertheless, the operation described allows us to cluster the whole picture quickly.

3 Description of the Algorithm

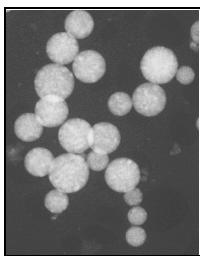
In this section, we present a detailed description of proposed algorithm. As was mentioned earlier, CEC works on a binarized version of images, therefore we begin by thresholding by the Otsu algorithm [8].

In a natural way, thresholding has a great influence on the final algorithm results. In some situations it causes disappearing of small circles (which are interpreted as elements of the background) or deforming of existing one (by adding background elements). As we mentioned earlier, our method copes well with such set distortion and not disjoint disks. Nevertheless, CEC gives better results if analyzed shapes are disjoint disks. Therefore, we use a morphology operation [10, 12], namely erosion. Thanks to this, we obtain splitting the dataset into small connected components which contains more separated elements. Erosion gives a good effect in the cases of a picture containing groups of circles which are not disjoint. We present this kind of a picture in Fig. 3. This modification divides the set and extracts circles. After erosion, the circles radius are smaller than this on the original picture. The resizing depends on a structural element used in erosion.

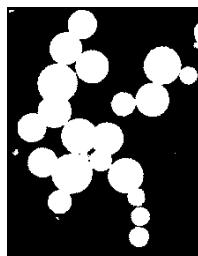
In the next step, we divide the picture into connected components. This allows us to work with smaller amounts of data. In many cases we obtain groups containing only a few disks, where CEC works fast sufficiently. After division we apply CEC to each component.



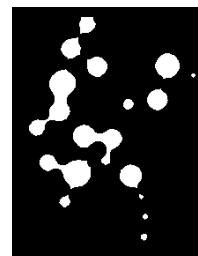
(a) Example of TEM picture.



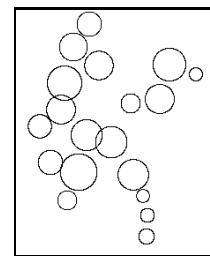
(b) The original dataset.



(c) The binarizatoin of the image.



(d) The image after erosion.



(e) Circles detected by the CEC algorithm.

Fig. 3 The results of the CEC algorithm with radius corrected respectively to the structural element which was used in erosion

Our method gives good results in most of the cases, nevertheless we are not able to detect circles that are not fully in the picture (see the left bottom corner or center bottom of 4b) and we loose some elements by thresholding (observe the disappearing of small circles in the upper left corner of 4a).

The final algorithm can be described as follows:

- Resize the picture (optionally).
- Use the Otsu algorithm for binarization of the picture
- Apply erosion filter with a circle–like structural element.
- Divide the set into connected components.
- Apply CEC in each cluster.
- Correct the radius of the circle obtained by CEC respectively to structural element which was used in erosion.

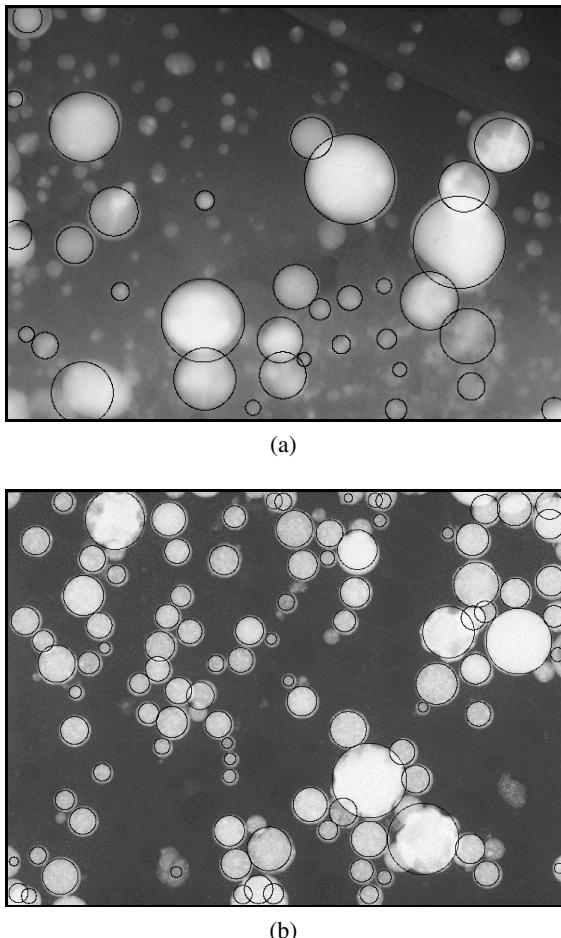


Fig. 4 The effect of the CEC algorithm

Our method works well in the case of resized pictures so fitting the parameters is quick. In general our method needs the following parameters:

- The level of thresholding (in this article we used the Otsu method).
- The size of a structural element in the erosion filter (we used approximately half of the radius of the smallest disks in a cluster)
- Two parameters requested by the CEC algorithm:
 - The starting number of clusters in each connected component (approximately the number of circles in the largest one)
 - The parameter describing the level of a minimal cluster size (we fixed it at 10% of the connected component size).

Fig. 4 presents the effect of our algorithm.

4 Conclusion

In this paper we present a new method of disc-like shapes detection in TEM pictures. As was mentioned earlier, in this situation we need to deal with elements which are not exactly disks, do not have smooth borders and in many cases they are not disjoint. Our approach, which is based on the CEC algorithm, allows us to deal with extracting this kind of elements.

The results of our investigation can be applied in the case of many types of nano-materials with nanoparticles with spherical shapes. Moreover we can use the method to general disc detection in various type of images.

References

1. Aguado, A.S., Nixon, M.: Feature Extraction & Image Processing for Computer Vision. Academic Press (2012)
2. Cover, T.M., Thomas, J.A., Wiley, J.: Elements of information theory, vol. 6. Wiley Online Library (1991)
3. Chan, W.C.W.: Bio-applications of Nanoparticles. Springer (2007)
4. Giersig, M., Khomutov, G.B.: Nanomaterials for application in medicine and biology. Springer (2008)
5. Hough, P.V.C.: Method and means for recognizing complex patterns. Google Patents, US Patent 3,069,654 (1962)
6. Ivakin, Y.D., Danchevskaya, M.N., Muravieva, G.P.: Regulation of gahnite crystal size during hydrothermal synthesis. Journal of Physics: Conference Series 8 (2008)
7. Illingworth, J., Kittler, J.: A survey of the Hough transform Computer vision, graphics, and image processing 44(1), 87–116 (1988)
8. Otsu, N.: A threshold selection method from gray-level histograms. Automatica 11, 23–27 (1975)
9. Salata, O.V.: Applications of nanoparticles in biology and medicine. Journal of Nanobiotechnology 2 (2004)
10. Shih, F.Y.: Image Processing and Mathematical Morphology: Fundamentals and Applications. CRC Press, Inc. (2009)
11. Tabor, J. Spurek, P.: Cross-entropy clustering (2012),
<http://arxiv.org/pdf/1210.5594.pdf>
12. Wilkinson, M.H.F., Roerdink, J.B.T.M.: Mathematical Morphology and Its Application to Signal and Image Processing. In: Wilkinson, M.H.F., Roerdink, J.B.T.M. (eds.) ISMM 2009. LNCS, vol. 5720, pp. 115–125. Springer, Heidelberg (2009)
13. Varadan, V.K., Chen, L., Xie, J.: Design and Applications of Magnetic Nanomaterials, Nanosensors and Nanosystems. Wiley Online Library (2008)
14. Yuen, H.K., Princen, J., Illingworth, J., Kittler, J.: Comparative study of Hough transform methods for circle finding. Image and Vision Computing 8(1), 71–77 (1990)

A GPU Accelerated Local Polynomial Approximation Algorithm for Efficient Denoising of MR Images

Artur Klepaczko

Abstract. This paper presents a parallelized implementation of the Local Polynomial Approximation algorithm targeted at CUDA-enabled GPU hardware. Although the application area of LPA in the image processing domain is very wide, here the focus is put on magnetic resonance image de-noising. In this case, LPA serves as a pre-processing step in the method based on Shape-Adaptive Discrete Cosine Transform. It is shown, how the designed efficient implementation of LPA substantially reduces the execution time of SA-DCT.

1 Introduction

The problem of noise removal from MR images has been extensively studied and variety of effective solutions were already proposed. It is an important problem while noise-free images can largely improve correctness of medical diagnosis based both on qualitative assessment as well as automatic computer-aided pattern recognition tools (segmentation, classification). Typically, the signal-to-noise ratio in real images results from the trade-off between acquisition speed, resolution, scanner field strength. The sources of noise in MR are multifold, including thermal phenomena, inductive losses, or sampling frequency.

A common assumption underlying majority of approaches derives from the observation that noise signal in magnitude MR images can be modeled by the Rician distribution [9]. Furthermore, it can be shown, that in the case of high SNR, distribution of noise approaches the Gaussian model. Therefore, the most straightforward techniques applied to MR image denoising are based on Gaussian or Wiener filters. Another approach involves restoration based on non-linear optimization of an image modeled as a Markov random field [4]. However, these methods usually fail leading to over-smoothed and blurred images, corrupted also by edge-related artifacts.

Artur Klepaczko
Lodz University of Technology, Institute of Electronics,
ul. Wolczanska 211/215, 90-924 Lodz
e-mail: aklepaczko@p.lodz.pl

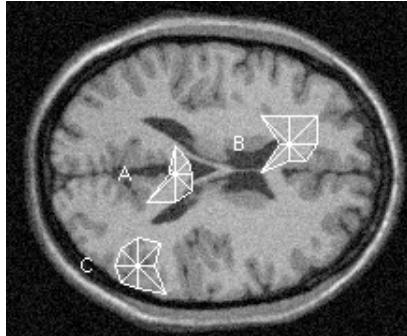


Fig. 1 Example pixels neighborhood masks determined by the LPA algorithm. Masks are adjusted to true local image contents while noise-associated intensity fluctuations are tolerated.

Therefore, the more sophisticated techniques either apply Rician model without its simplification (e.g. joint geometric-, radiometric-, and median-metric filter [2], Median Absolute Deviation estimator or Nonlocal Maximum Likelihood [3,5]) or—although assume the noise to be normally distributed—operate on voxels neighborhoods adapted to fine details of visualized structures, thus making noise removal more robust to blurring and over-smoothing problems. In this paper we focus on one such method, namely the Shape-Adaptive Discrete Cosine Transform (SA-DCT) [6]. As a first step, SA-DCT determines homogeneous local neighborhoods of every image voxel using Local Polynomial Approximation algorithm. The neighborhoods are finely adjusted to shape and size of local image contents (cf. Fig. 1). However, determination of local neighborhoods for every voxel involves significant computational overhead. Application of LPA to high-resolution 3D MR images occurs inefficient. Thus, in this paper parallelized GPU-accelerated implementation of LPA is proposed, so that noise can be removed from MR data not only effectively, but also efficiently.

2 Local Polynomial Approximation

The LPA algorithm is a technique of non-parametric regression adopted in various image processing applications [1, 7]. Using low order polynomial function, LPA models a non-linear relationship between an independent variable X and a dependent variable Y . Data are fitted to a modeled polynomial function within a sliding window positioned at subsequent observations (X, Y) – e.g. measured values of a sampled signal. In a window, a signal is convolved with a kernel function of a known form. This enables estimating values of Y in the neighborhood of a given data point X . Window size h is a key parameter of the method. It is defined as a number of data samples beyond which it becomes impossible to estimate signal Y basing on values measured in the proximal neighborhood of X .

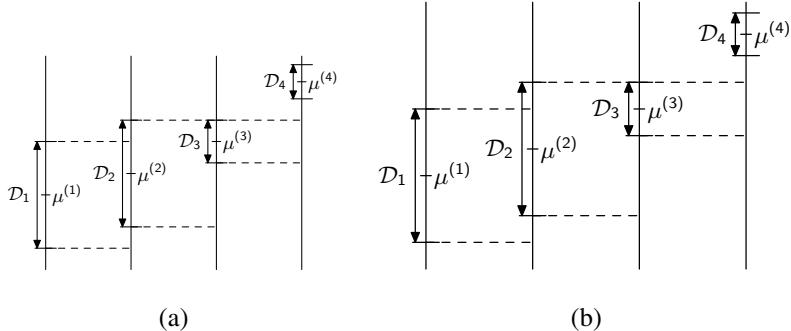


Fig. 2 Distinct directions $\theta_i, i = 1..8$ used in LPA filtering (a) and the illustration of the Intersection of Confidence Intervals rule for $h_{\max} = 3$ (b).

In this study the 3D image is considered as a stack of flat slices and thus a 2D variant of LPA algorithm is applied. Then, each pixel neighborhood is filtered in 8 distinct directions, as illustrated in Fig. 2a. For a given pixel X , the filtered value μ is calculated as

$$\mu^{(h)} = \sum_{j=1}^h g_j^{(h)} I(X + (j-1)\theta_i), \quad (1)$$

where $g^{(h)}$ is a discrete convolution kernel of scale h (window size), $g_j^{(h)}$ with $j = 1, \dots, h$ denote kernel weights which sum to unity and decrease with the increasing distance from a center pixel X . The exact procedure of weights generation is described in [7]. I is a 2D matrix of image intensity values.

Adjusting the window size to local image contents is performed using the *intersection of confidence intervals* (ICI) rule. The idea is to test several values of scale h , i.e. $h \in \{h_1, \dots, h_k\}$ and $h_1 < h_2 < \dots < h_k$ and for each of them evaluate (1) as well as local standard deviation value

$$\sigma_{\mu^{(h)}} = \sigma \|g^{(h)}\|, \quad (2)$$

where σ is the global standard deviation determined for the whole image. Then for each direction θ_i and scale h one calculates confidence intervals

$$\mathcal{D}_h = [\mu^{(h)} - \Gamma \sigma_{\mu^{(h)}}, \mu^{(h)} + \Gamma \sigma_{\mu^{(h)}},] \quad (3)$$

in which $\Gamma > 0$ denotes a global parameter that allows controlling noise tolerance. The lower Γ , the stronger requirement for local homogeneity is, and thus fewer pixels are included in the resulting neighborhood volumes. The ICI rule states that for each direction one should choose a maximum value of h that ensures nonempty intersection of all previous confidence intervals, i.e. (cf. Fig. 2b)

$$h_{\max,i} = \max_{h \in \{h_1, \dots, h_k\}} \{h : (\mathcal{D}_1 \cap \mathcal{D}_2 \cap \dots \cap \mathcal{D}_h) \neq \emptyset\}. \quad (4)$$

In this study it is arbitrarily set $h \in \{1, 2, 3, 5, 7, 8\}$, hence the upper bound for the window size in any direction amounts to 8 pixels. On completion, pixels determined by relations $X + h_{\max,i} \theta_i$ constitute a set of a hull vertices whose interior determines a locally adapted volume of interest of X .

3 GPU-Based Implementation of LPA

One of the most important features of the CUDA technology is the separation of the code layer from the GPU hardware, allowing seamless execution of the same binaries on different devices. However there are some hardware-related details specific to various generations of Nvidia's products and thus it should be noted, that this study targets the *Fermi architecture* (GF116 in particular). Below the designed implementation is presented in 3 variations, starting from the most straightforward solution. Then, additional optimization mechanisms are introduced to improve ultimate efficiency. The presentation is preceded by recalling basic configuration issues common to any CUDA project.

3.1 The Execution Configuration

The CUDA programming model assumes problem decomposition into a series of threads, each executing the same operation—called a *kernel*—on different portions of data. CUDA threads have to be organized in blocks, and then blocks of threads build up a grid. The total number of threads has to comply with the problem size, i.e. every data element (an image pixel) is assigned its own corresponding thread. The grid organization into blocks of threads is referred to as *execution configuration* and it is important to properly adjust it in order to achieve the maximal utilization of GPU hardware resources. All CUDA-enabled processors are composed of the Streaming Multiprocessors (SM). In the Fermi architectures a SM can be assigned at most 8 blocks or 1536 threads at a time. Therefore, too large (e.g. 32×32 threads) or too small (e.g. 8×8) blocks results in poorer utilization of computational logic and potential loss in performance. Percentage of threads actually assigned to a SM in relation to maximum possible number of threads per SM is called the *occupancy ratio*.

Beside execution configuration, there are other factors which may cause occupancy to be lower than optimal 100%. These relate mainly to memory resources. In CUDA there are four major types of memory. The fastest accessible are per-thread registers. They are used mainly for storing scalar variables declared in a kernel scope. In GF116 there is a limit of 32K of 32-bit registers per multiprocessor. Thus, if a SM is assigned a total number of 1536 threads, no more than 21 registers are available for a thread. It may however be beneficial to speed up kernel execution by extensive usage of registers at the cost of occupancy. Secondly, each thread can access its own local memory area (of 512 KB size), which is however a

long-latency, off-chip storage. The third memory option is 48KB of shared memory to be distributed among blocks in a SM. Similarly to registers, it is zero-overhead memory but its over-utilization by a block may result in degraded occupancy. Eventually, global memory is a large capacity data storage, available to all threads in a grid throughout the whole application lifetime. However, it is again, an off-chip long latency storage.

Taking into account the above considerations, the following configuration is used in the proposed design. The block size is set to 16×16 . This gives 256 threads per block and allows 6 blocks to simultaneously reside in SM. The grid size is adjusted to match the processed image. For an image of size 256×256 pixels, the grid dimension will be 16×16 . In this configuration, the amount of shared memory per block is limited to 8KB (if 100% occupancy is to be maintained). The number of registers per kernel is kept equal to 21 (it can be fixed at compile time), however it could be tuned to improve final performance.

3.2 The Basic Kernel

The first step of LPA is to estimate global standard deviation σ of noise in an input image. This can be accomplished relatively fast in a single CPU thread. Then, the CPU part of the program (the *host*) transfers image data to the the GPU global memory and executes kernel in a grid of threads, where every thread evaluates equations (1) to (4) in its own dedicated pixel X . Thus, the kernel has to begin with identifying its pixel coordinates and proceeds as shown in Fig. 3. Note, that in this basic implementation the thread must read global memory in step 4 at least 9 times—to fetch its pixel and the nearest neighbors intensity values. In the worst case, where the ICI rule holds for the most distal pixels in all 8 directions, the number of global memory transactions in step 4 reaches value of 65 which can cause significant computational overhead.

3.3 Shared Memory

Shared memory can be used to reduce the extensive traffic between GPU and off-chip global memory, in which many transfers concern the same data. Note that adjacent pixel neighborhoods investigate partially the same image region. For example, two pixels which differ only on horizontal coordinate by one require—in direction $\theta_1 = 0^\circ$ —analysis of an image row whose size is 10 pixels long, 8 of which must be read by both GPU threads in the basic kernel (see Fig. 4a). For the whole 16×16 threads block a common image region is 32×32 pixels large. This region can be efficiently loaded in 4 stages into shared memory space (cf. Fig. 4b).

In every stage, a subregion of size 16×16 pixels is retrieved. The subregions are shifted relative to an image region associated to a thread-block. For example, in the case of subregion I, a thread reads a pixel intensity which is located 8 points

1. Determine pixel coordinates based on thread and block ID numbers and dimensions.
2. For each direction $\theta_i, i = 1,..,8$
3. For each scale $h_j \in h_1,..h_k$
 4. Calculate μ^h according to (1)
 5. Calculate $\sigma_{\mu^{(h)}}$ according to (2)
 6. Calculate \mathcal{D}_{h_j} according to (3)
 7. If the ICI rule (eq.(4)) holds for \mathcal{D}_{h_j}
 8. go to step 4.
 9. else break
 10. Store scale h_j^i for direction θ_i and go to step 2.

Fig. 3 Pseudo code of the basic CUDA implementation of the LPA algorithm

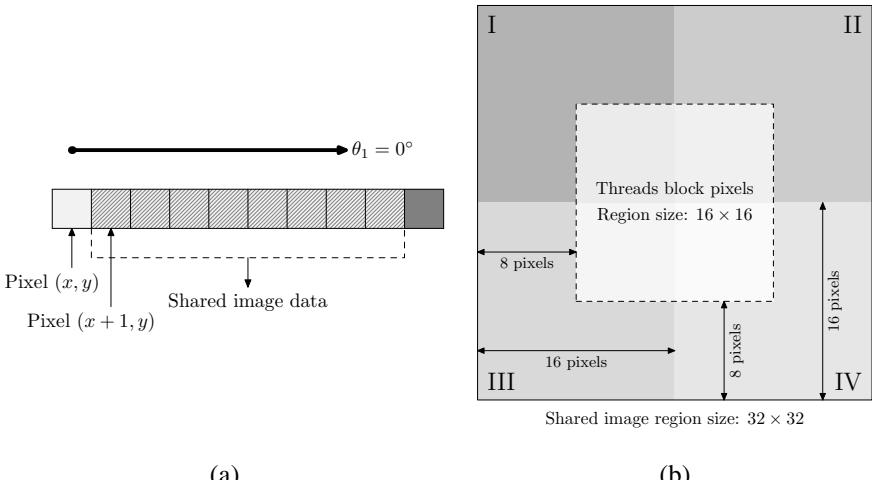


Fig. 4 Shared image data for two adjacent pixels in direction $\theta_1 = 0^\circ$ (a) and for the whole thread block (b)

to the left and 8 points above this thread dedicated pixel. After requests for the last subregion data are sent to global memory, there must be a synchronization barrier set so that transfers scheduled by all threads in a block complete before computations begin. Finally, a kernel proceeds from step 2 in the algorithm listed in Fig. 3. This time however, data requests induced in step 4 refer to shared and not global memory. In this way, there are only 4 instead of maximally 65 global memory reads.

Note, that the amount of shared memory that must be allocated for the region of size 32×32 is 4kB if the image pixels are described by 32-bit floating point

numbers (the usual data type of MRI images). This volume fits the limit of 8kB per block determined by the execution configuration.

3.4 Increased Thread Responsibility

Additional mechanism allowing to improve kernel performance is to re-use the data once they are loaded from global memory. One can notice that an image region loaded into shared memory as shown in Fig. 4 already contains a substantial portion of information needed to calculate LPA masks for one of the neighboring thread blocks. Thus, to make profit of data re-use, every thread can be made responsible for two image pixels. This requires allocation of 6kB of shared memory space covering an image region of size 32×48 pixels. Hence, there are only 2 additional global memory reads (50% more) for a thread while the number of LPA masks calculated in a kernel doubles. The execution time of a single thread grows, but the number of threads that needs to be invoked reduces by half and this leads to considerable performance boost as reported in the next section.

4 Experiments

Efficiency of the proposed implementation was tested in a series of experiments performed on 10 2D simulated brain MR magnitude images [8]. Width and height of 2D slices—originally 181×217 pixels—were zero-padded to match the size of 256×256 pixels. One half of images was degraded by additive Gaussian noise of zero mean and variable standard deviation, i.e. $\sigma = 0.01, 0.03, 0.05, 0.07$ and 0.15 . The other half was corrupted with the Rician noise modeled as

$$p(m|X) = \frac{m}{\sigma^2} \exp^{-(m^2 + X^2)/2\sigma^2} I_0\left(\frac{Xm}{\sigma^2}\right) \quad (5)$$

where m denotes corrupted image pixel, X is the noise-free intensity of the pixel, and σ is the standard deviation of the underlying normally-distributed noise, which—in real conditions—adds to raw complex MR data. After Fourier transform of k-space, these data become Rician-distributed. Thus, in order to obtain noisy image from simulated brain phantom the following equation was applied to every image pixel

$$m = \sqrt{A^2 + B^2}, \quad (6)$$

where $A \sim N(X \cdot \cos(a), s^2)$ and $B \sim N(X \cdot \sin(a), s^2)$ are independent normal distributions (any real a). The parameter s can be treated as *noise level*, which in the conducted experiments was set to $s = \{1, 2, 3, 4, 5\}$. Measurements presented below are the average estimates obtained for all tested images. Fig. 5 presents example 2D images corrupted with Gaussian and Rician noise along with a sample result of

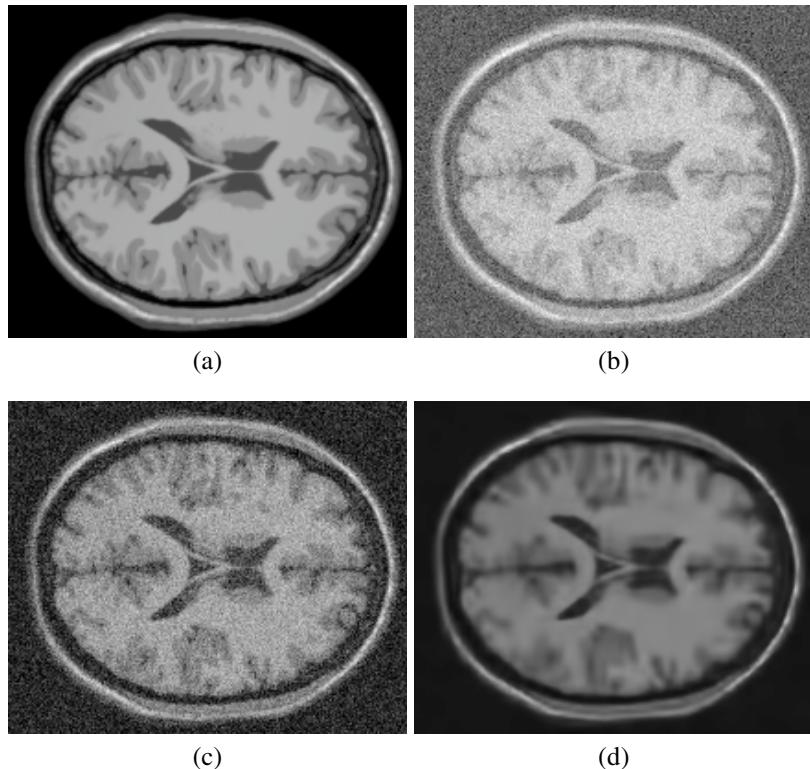


Fig. 5 Example 2D slice of synthetic brain image: (a) noise-free, (b) corrupted with Gaussian noise ($\sigma = 0.15$), (c) corrupted with Rician noise ($s = 5$), (d) denoised image (b)

Table 1 Execution times of LPA kernels [ms] under various implementation designs

	CPU	Basic	Shared	Shared×2
Host-to-device data transfer	N/A		0.81	
LPA kernel	721.0	2.34	2.03	1.73
Device-to-host data transfer	N/A		4.53	

noise-removal procedure accomplished using the designed implementation of LPA algorithm as part of the SA-DCT method.

The GPU code was run on the GF116-compliant GeForce GTX 560M chip. Since the graphics processor used in the experiments is targeted at mobile devices, also the CPU chip chosen for tests was a mobile variant of the Intel Core i7 (i7-2630QM). Time records viewed in Table 1 were measured using the CUDA Event API.

Table 2 Execution times of LPA kernels vs. number of registers used by a thread

No. of registers	21	23	25	27	29	31
Execution time [ms]	1.73	1.68	1.65	1.63	1.62	1.65
Occupancy [%]	100	92	85	79	73	69

5 Results Discussion and Conclusions

Analysis of the obtained results shows superior performance of the GPU accelerated implementation of LPA over an analogous program run on CPU. In the latter case though, the time was measured for a single CPU thread. However, even if the score was divided by a factor of 8 (theoretical number of threads which can be simultaneously executed on i7 processor), GPU code runs on average and depending on the implementation variant 38 to 52 times faster. These ratios scale down to 11.7 and 12.7 if host-to-device and device-to-host data transfers are taken into account.

Moreover, it can be noticed how usage of shared memory speeds up computations. Execution time in the implementation variant described in Sect. 3.3 is 20% lower than in the case of the basic kernel. Increased responsibility of a kernel (this variant is denoted 'Shared×2' in Table 1) leads to even higher performance. Eventually, to test how the number of registers used by a thread affects the overall efficiency, the program (version 'Shared×2') was compiled using variable option `maxreg` in the CUDA `nvcc` compiler. Recall that excessive usage of registers—although may speed up a kernel execution—degrades the occupancy. However, as shown in Table 2, despite lower occupancy the optimum for the designed implementation is 29 registers per thread. Increasing the number of registers from 21 to 29 leads to observable improved efficiency. This trend halts only after the occupancy drops below 70%.

To conclude, it must be underlined that the designed GPU-based implementation of the LPA algorithm performs very efficiently. Accomplished within a timeframe reduced to miliseconds, generation of LPA masks no longer entails any significant computational load to SA-DCT-based noise removal method.

Acknowledgements. This paper was supported by the Polish National Science Centre grant no. N N519 650940.

References

1. Bergmann, Ø., Christiansen, O., Lie, J., Lundervold, A.: *J. Digital Imaging* 22(3), 297–308 (2009)
2. Chang, H.H.: Rician noise removal in MR images using an adaptive trilateral filter. In: 2011 4th International Conference on Proc. Biomedical Engineering and Informatics (BMEI), pp. 467–471 (2011)

3. Coupé, P., Manjón, J., Gedamu, E., Arnold, D., Robles, M., Collins, D.L.: Medical Image Analysis 14(4), 483–493 (2010)
4. Garnier, S.J., Bilbro, G.L., Snyder, W.E., Gault, J.W.: J. Digit. Imaging 7(4), 183–188 (1994)
5. He, L., Greenshields, I.R.: IEEE Trans. Medical Imaging 28(2), 165–172 (2009)
6. Katkovnik, V., Egiazarian, K., Astola, J.: J. Math. Imaging and Vision 16, 223–235 (2002)
7. Katkovnik, V., Egiazarian, K., Astola, J.: Local Approximation Techniques in Signal and Image Processing. SPIE Press (2006)
8. Kwan, R.S., Evans, A., Pike, G.: IEEE Trans. Medical Imaging 11, 1085–1097 (1999)
9. Sijbers, J., Dekker, J.d., Audekerke, J.V., Verhoye, M., Dyck, D.V.: Magnetic Resonance Imaging 16(1), 87–90 (1998)

Altair: Automatic Image Analyzer to Assess Retinal Vessel Caliber

Gabino Verde, Luis García-Ortiz, Sara Rodríguez,
José I. Recio-Rodríguez, Juan F. De Paz, Manuel A. Gómez-Marcos,
Miguel A. Merchán, and Juan M. Corchado

Abstract. The scope of this work is to develop a technological platform specialized in assessing retinal vessel caliber and describing the relationship of the results obtained to cardiovascular risk. Population studies conducted have found retinal vessel caliber to be related to the risk of hypertension, left ventricular hypertrophy, metabolic syndrome, stroke, and coronary artery disease. The vascular system in the human retina has a unique property: it is easily observed in its natural living state in the human retina by the use of a retinal camera. Retinal circulation is an area of active research by numerous groups, and there is general experimental agreement on the analysis of the patterns of the retinal blood vessels in the normal human retina. The development of automated tools designed to improve performance and decrease interobserver variability, therefore, appears necessary.

1 Introduction and Background

Image processing, analysis and computer vision techniques are increasing in prominence in all fields of medical science, and are especially pertinent to modern ophthalmology, which is heavily dependent on visually oriented signs. Automatic detection of parameters from retinal images is an important problem since are associated with the risk of hypertension, left ventricular hypertrophy, metabolic syndrome, stroke, and coronary artery disease [14] [15].

The vascular system in the human retina has a unique property: it is easily observed in its natural living state in the human retina by the use of a retinal camera.

Gabino Verde · Sara Rodríguez · Juan F. De Paz · Juan M. Corchado
Computers and Automation Department, University of Salamanca, Salamanca, Spain

Luis García-Ortiz · José I. Recio-Rodríguez · Manuel A. Gómez-Marcos ·
Miguel A. Merchán
Primary care Research unit La Alamedilla. Sacyl. IBSAL. Salamanca, Spain
e-mail: {gaby, lgarciao, srg, donrecio, fcofds, magomez}@usal.es,
merchan, corchado@usal.es

The retina is the only human location where blood vessels can be directly visualized non-invasively. The identification of landmark features such as the optic disc, fovea and the retinal vessels as reference co-ordinates is a prerequisite to systems being able to achieve more complex tasks that identify pathological entities. Reliable techniques exist for identifying these structures in retinal photographs. The most studied areas in this field can be classified into three groups [10]:

1. The *location of the optic disc*, which is important in retinal image analysis for vessel tracking, as a reference length for measuring distances in retinal images, and for identifying changes within the optic disc region due to disease. Techniques such as analysis of intensity pixels with a high grey-scale value [8] [3] or principal component analysis (PCA) [9] are used for locating the disk. Other authors [7] use the Hough transform (a general technique for identifying the locations and orientations of certain types of shapes within a digital image [7]) to locate the optic disc. A “fuzzy convergence” algorithm is another technique used for this goal [4].
2. The *detection of the fovea*, usually chosen as the position of maximum correlation between a model template and the intensity image [9].
3. The segmentation of the vasculature from retinal images, that is, the representation of the blood vessels and their connections by segments or similar structures. There are a lot of techniques to do this, the most significant of which are: (i) matched filters, which typically have a Gaussian or a Gaussian derivative profile; (ii) vessel tracking, whereby vessel centre locations are automatically sought over each cross-section of a vessel along the vessel's longitudinal axis, having been given a starting and end point [13]; (iii) neural networks, which employ mathematical “weights” to decide the probability of input data belonging to a particular output [1]; (iv) morphological processing, which uses characteristics of the vasculature shape that are known a priori, such as being piecewise linear and connected [4].

An understanding of the design principles of the human vascular system may have applications in the synthetic design of vascular systems in tissue and organ engineering, i.e., bioartificial organs for both liver and kidney. In current scientific literature one can find a lot of research devoted to automating the analysis of retinal images [12] [2] [4]. In this paper, after several years of studies and tests, we propose a novel platform image processing to study the structural properties of vessels, arteries and veins that are observed with a red-free fundus camera in the normal human eye, and the fractal analysis of the branching trees of the vascular system. The platform, called Altair "Automatic image analyzer to assess retinal vessel caliber", employs analytical methods and AI (Artificial Intelligence) algorithms to detect retinal parameters of interest. The sequence of algorithms represents a new methodology to determine the properties of retinal veins and arteries. The platform does not require user initialization, it is robust to the changes in the appearance of retinal fundus images typically encountered in clinical environments, and is intended as a unified tool to link all the methods needed to automate all processes of measurement on the retinas. The platform uses the latest computer techniques both statistical and medical. The next section introduces the Altair platform. Section 3 presents the most important characteristics of the platform, showing some of the relevant techniques and results. Finally, some conclusions are presented in section 4.

2 Platform Overview

Altair facilitates the study of structural properties of vessels, arteries and veins that are observed with a red-free fundus camera in the normal human eye, and the fractal analysis of the branching trees of the vascular system. Figure 1 shows an example of images taken directly from the fundus. The retinal vessels appear in a different color, with the optic disc and fovea. There are many patterns in nature that show branching, such as the retinal vessel, and those with open branching structures and different lengths. These objects can be described by fractal geometry. Different analytical methods and AI algorithms are used to determine the scaling properties of real objects, yielding different measures of the fractal dimension, length and area of retinal veins and arteries.

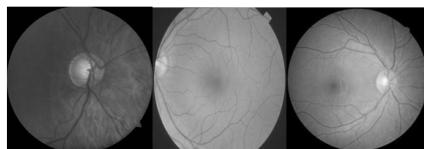


Fig. 1 Retinograph usually takes three images of each eye: a) Centered papilla. b) With the disc on one side. c) With the macula and disc each to one side of center

The main objective is to relate the level of cardiovascular risk in patients to everything that can be observed in the retinas. In this work we are interested in obtaining as much information as possible from the images obtained, and have focused on the following:

- Index Artery / Vein: represents a relationship between the thickness of arteries and veins.
- Branching: branching structures include fractal analysis of the branching trees of the vascular system. More branches tend to appear in subjects with cardiovascular diseases, especially around the papilla. Branching index refers to the number of times that an artery branches, while branching pattern refers to the way in which arteries branch. Actually the manner in which branching occurs is practically a fingerprint in that each person has a different shape, however in many samples retinas it is possible to observe certain normal patterns. In that case the relationship of these patterns with the diseases could be studied.

- Area occupied by the veins and arteries.
- Distribution of the capillary: according to the blood distribution, the color distribution of the capillaries varies.

Moreover, our intention is to incorporate expert knowledge taken from the measurements found in the retinal circulation which specify normal values of various retinal structures in healthy subjects, and to apply this information to the study of patients suffering from a number of diseases. Based on the values for area, length, position and patterns of the branching trees of the vascular system in healthy pa-

tients, we expect to determine ranges of normalcy within the population for their subsequent application to subjects affected by various diseases. The next section explains the main components of the platform. The original image passes through each one of the modules (preprocessing, detection, segmentation and extraction of knowledge), which use different techniques and algorithms to obtain the desired image information. This sequence of steps is a methodology that is explained in the following section, also showing examples of the results obtained.

3 Methodology and Results

The methodology used to obtain the functionality of the platform may be divided into two phases. Firstly, a phase called "digitization of the retina", in which the different parts of the eye image are identified. Here a data structure is created, which makes it possible to represent and process the retina without requiring the original image. This phase includes modules of preprocessing, detection and segmentation. Secondly, a phase of "measurements" in which we work with retinas that have been previously identified. This phase includes extraction of knowledge and manual correction, or expert knowledge, if necessary.

This paper focuses on the first phase, which is in charge of creating and identifying all the elements of interest of the retina. To carry out these phases, the following steps are necessary.

Preprocessing

The preprocessing or filtering module reduces noise, improves contrast, sharpens edges or corrects blurriness. Some of these actions can be carried out at the hardware level, which is to say with the features included with the camera. During the testing, retinography was performed using a Topcon TRC NW 200 nonmydriatic retinal camera (Topcon Europe B.C., Capelle a/d IJssel, The Netherlands), obtaining nasal and temporal images centered on the disk (Figure 1). The nasal image with the centered disk is loaded into the platform software through the preprocessing module.

Detection limits

This module is in charge of locating the disk and identifying the center and edges of the retina. The goal here is to construct a data structure that identifies each part of the retina based on the matrices of colors representing the images obtained (Figure 1). In this step, image processing techniques were used to detect intensity based on the boundaries of the structures [4] [2]. The identification of the papilla is important since it serves as the starting point for the detection and identification of the different blood vessels. This phase identifies the boundaries and the retinal papilla from a RGB image of the retina. The following values are returned: Cr is the center of the retina, which identifies the vector with coordinates x, y of the center of the retina. Cp is the center of the disc, which identifies the vector with the coordinates x, y

of the center of the papilla. R_r , is the radius of the retina. R_p , is the radius of the papilla. As an example, a sequence of output values in this phase is shown in the following table and figure:

C_r	C_p	R_r	R_p
1012,44 ; 774,13	1035,98 ; 734,11 1104,87 ; 562,52 915,38 ; 736,77 900,27 ; 658,74	692,68	111,76 108,92 122,15 101,95



Fig. 2 Identification result in the detection phase. Table 1. equence of output values in detection modules (pixel).

In order to identify the limits, and in particular to identify the circumferences, it became necessary to carry out a process of image segmentation. Segmentation is the process that divides an image into regions or objects whose pixels have similar attributes. Each segmented region typically has a physical significance within the image. It is one of the most important processes in an automated vision system because it makes it possible to extract the objects from the image for subsequent description and recognition. Segmentation techniques can be grouped into three main groups: techniques based on the detection of edges or borders [7], thresholding techniques [8], and techniques based on clustering of pixels [3]. After analyzing the possibilities, we chose one of the techniques from the first group that provided the best results, which in this case uses an optimization of the Hough transform [7]. This technique is very robust against noise and the existence of gaps in the border of the object. It is used to detect different shapes in digital images. When applying the Hough transform to an image, it is first necessary to obtain a binary image of the pixels that form part of the limits of the object (applying edge detection). The aim of the Hough transform is to find aligned points that may exist in the image to form a desired shape. For example, to identify line points that satisfy the equation of the line: $\rho = x \bullet \cos\theta + \sin\theta$, in polar coordinate. In our case, we looked for points that verify the equation of the circle: (i) in polar coordinate system: $r^2 - 2sr \bullet \cos(\theta - \alpha) + s^2 = c^2$, where (s, α) is the center and c the radius; (ii) in Cartesian coordinate system: $(x - a)^2 + (y - b)^2 = r^2$, where (a, b) is the center and r the radius. The algorithm is not computationally heavy, as it does not check all radios, or all possible centers, only the candidate values. The candidate centers are those defined in a near portion of the retina, and the radius is approximately one sixth the radius of the retina. To measure the approximate diameter of the retina, the algorithm calculates the average color of the image column: diameter of the retina is the length that has a non-zero value (black).

Identifying the papilla (Figure 3) is a necessary step because it provides a starting point for other stages of segmentation and serves as a reference point for some typical measurements. Typically the correct result is the circumference of the higher value in the accumulator (over 70% of cases). In almost 100% of the cases, the cor-

```

load image
detect edges
for each candidate point (a, b) in the image
    for each candidate radius r
        calculate the points (x, y) which are at an edge and
        y is in the circumference of center (a, b) and radius r and
        introduce them into the accumulator
find the pairs (center, radius) whose values are the highest in the accumulator

```

Fig. 3 Pseudocode of the identification algorithm of the papilla

rect identification can be found among the 3 greatest values found by the accumulator.

Segmentation of the vasculature from retinal images

The ultimate goal in this module is to identify each blood vessel as a series of points that define the path of the vessel. Each of these points will be assigned a certain thickness. Moreover, it will be necessary to distinguish whether a particular blood vessel is a vein or an artery. AI algorithms responsible for identifying veins and arteries must perform a series of sweeps in search of "key points". Algorithms based on matched filters [6], vessel tracking [12] and PCA [9], among others, are used for obtaining the proximity points between objects (veins, arteries, capillaries), the structures retinal structures or assemblies, branching patterns, etc. These algorithms work with transformations of the original image of the retina obtained from the previous step. Three steps are necessary within this module: (i) identification of vessels; (ii) definition of the structure of vessel; (iii) cataloging of veins and arteries.

Identification of vessels. In this step the blood vessels are identified in the image by thresholding techniques. Their purpose is to remove pixels where the structuring element does not enter, in this case the blood vessels. The image on the retina is blurred to keep an image similar to the background. This image is used as a threshold so that the pixels of the original image will be treated as vessels if their intensity reaches 90% of the background intensity.

The image below represents the application of these techniques in a row. The blue line represents the values of the pixels in the image; the red line, the background values; and the green line the point where there is a vessel:



Fig. 4 Thresholding techniques for the identification of vessels

In the figure, it is possible to observe a very small vessel on the left, which comes from the artery below. In the middle there is a fat vein and to the right there are three tiny vessels. Furthermore the edge of the retina is marked as a vessel although obviously it is not. To decide where there is a vessel, the following algorithm is

applied (Figure 5a) where $Original(x,y)$ is the pixel (x,y) of the original image and $Background(x,y)$ is the pixel (x,y) of the background image. The result is shown in Figure 5b.

```
If (Original (x, y) < Background (x, y) * threshold)
    There is vessel at (x, y)
If not
    There is not vessel at (x, y)
Threshold ~0.9
```



Fig. 5 a) Pseudocode of the identification algorithm of the papilla. b) Image result. (Blank pixels are the vessels).

Structure of vessel. This phase defines the tree forming blood vessels. Various techniques are used in conjunction with the following steps:

The following image shows the output of this phase. At the end of this stage the entire arterio-venous tree is stored in a structured way, making it possible to know not only if a vessel passes through a point or not, but through which point each vessel passes, which one is its parent, etc.

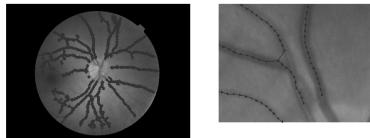


Fig. 6 Structure of the vessels

Cataloging of veins and arteries. To detect whether a vessel is vein or artery, main branch is taken of the vessel. For every point (x,y) of the branch:

```
If (Original (x,y) < Background (x, y) * threshold)
    Probable vein
If not
    Not vein
Threshold ~0.7
```

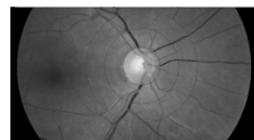


Fig. 7 a)Pseudocode of the identification algorithm of veins. b)Arteries and veins detection.

In general, if most points of the main branch of the vessel (from at least 60%) are points classified as "probable vein", we conclude that this vessel is a vein, otherwise an artery. Currently, there are no publicly available databases that can be used to assess the performance of automatic detection algorithms on retinal images. In this

study, we assessed the performance of our platform using retinal images acquired from the Primary Care Research Unit La Alamedilla, SACYL, IBSAL, Salamanca, Spain. The images were obtained using a TopCon TRC-NW6S Non-Mydriatic Retinal Camera. Table 2 shows the testing performed using 10 retinal images. No difference was found between values in terms of age, sex, cardiovascular risk factors, or drug use. The first row of values is shown in the examples retina and previous figures in this paper. The table shows: Area veins and arteries, Diameter of veins and arteries (D), AV index (AV), Veins P (VP) = number of veins around the papilla, Veins A (VA)= number of veins that cross the corona outlined with radius=2*Rp. Rp is the radio of the papilla, Veins B (VB)= number of veins that cross the corona outlined with radius=3*Rp, same values for arteries. And the ratios leaving the region around the papilla and out of the disc (which could serve as a reference of bifurcations that have occurred, though not in the manner in which they branch).

Ret	Area veins (mm ²)	Area arteries (mm ²)	R _p (mm)	V _P	V _A	V _B	A _P	A _A	A _B	Veins Thickness (mm)	Arteries Thickness (mm)	AV Index	Veins B/P	Veins B/B
1	2.200147433	1.997102803	1.13693	5	6	8	7	9	12	0.12084217	0.083101875	0.68768936	1.6	1.7142857
2	2.004176485	1.995543005	1.04058	4	5	6	5	8	11	0.15342774	0.098325175	0.64085657	1.5	2.2
3	1.637113923	1.377180893	0.973135	5	6	6	4	6	10	0.12165151	0.08752434	0.7194776	1.2	2.5
4	1.698848018	1.463144459	1.05985	5	6	8	6	9	10	0.12091925	0.08029809	0.66406375	1.6	1.6666667
5	2.047808111	1.156423484	1.0357625	6	6	10	6	6	8	0.112816215	0.098999625	0.87753011	1.66666667	1.33333333
6	1.845338847	1.641012918	1.07912	5	5	7	7	8	11	0.11180454	0.083313845	0.74517408	1.4	1.5714286
7	1.737095306	1.240159053	1.0800835	5	7	8	8	8	11	0.12290406	0.07945021	0.64644089	1.6	1.375
8	1.612513118	1.261882027	1.0396165	5	6	8	7	10	12	0.12195983	0.08199385	0.6723021	1.6	1.7142857
9	1.566189339	1.183159453	0.9008725	6	7	8	6	7	7	0.100271445	0.094124315	0.93869511	1.33333333	1.16666667
10	1.609078289	1.13962067	0.934595	6	7	8	5	6	9	0.13317497	0.099808945	0.74945739	1.33333333	1.8

Fig. 8 Output results for 10 retinal images. Table 2.

It is possible to observe that the measurement values of veins and arteries (thickness, area) are similar between different retinas (in this case no retinal images of sick patients were introduced). Parameters like the veins in the papilla and AV index are the most fluctuating. Due to the lack of a common database and a reliable way to measure performance, it is difficult to compare our platform to those previously reported in literature. Although some authors report algorithms and methods [12] [2] [4] with similar performance to our platform, these results may not be comparable, since these methods are tested separately and were assessed using different databases. Since automation has been valid and verified our next step is to compare the values obtained with significant medical values in our database.

Knowledge extraction

This platform will show a high intra-observer and inter-observer reliability with the possibility of expert corrections if necessary. Results of its validity analysis must be consistent with the findings from large studies conducted with regards to both cardiovascular risk estimation and evaluation of target organ damage. The results obtained during the use of the platform will be connected and used to extract additional information by using reasoning models such as case-based reasoning (CBR) [11].

Taking into account the measures found in the retinal circulation which specify normal values of various retinal structures in healthy subjects, it is possible to apply this information to the study of patients suffering from diseases. Moreover, because of the platform's semi-automated nature and rapid assessment of retinal vessels, it may be helpful in clinical practice.

4 Conclusions

Platforms such as Altair, which allow the automated diagnosis of retinal fundal images using digital image analysis, offer a lot of benefits. In a research context, they offer the potential to examine a large number of images with time and cost savings and offer more objective measurements than current observer-driven techniques. Advantages in a clinical context include the potential to perform large numbers of automated screening for conditions such as risk of hypertension, left ventricular hypertrophy, metabolic syndrome, stroke, and coronary artery disease, which in turn reduces the workload required from medical staff. As a future line of study in this point, the next step would be to analyze the significance of the measurements obtained with regard to their meaning in a medical context. That is, to describe the relationship of the results obtained to the risk of cardiovascular disease estimated with the Framingham or similar scale and markers of cardiovascular target organ damage. The platform is intended as a unified tool to link all the methods needed to automate all processes of measurement on the retinas. It uses the latest computer techniques both statistical and medical. Thanks to the experience of the research group at the University of Salamanca (bisite.usal.es), another line of the future, which is already underway, is the migration of the platform to a cloud environment in which all services are accessible by users, regardless of their location, both safely and ubiquitously [5].

References

1. Akita, K., Kuga, H.: A computer method of understanding ocular fundus images. *Pattern Recogn.* 16, 431–443 (1982)
2. Chen, B., Tosha, C., Gorin, M.B., Nusinowitz, S.: Analysis of Autofluorescent retinal images and measurement of atrophic lesion growth in Stargardt disease. *Experimental Eye Research* 91(2), 143–152 (2010)
3. Goldbaum, M., Katz, N., Nelson, M., Haff, L.: The discrimination of similarly colored objects in computer images of the ocular fundus. *Invest. Ophthalmol. Vis. Sci.* 31, 617–623 (1990)
4. Heneghan, C., Flynn, J., O'Keefe, M., Cahill, M.: Characterization of changes in blood vessel and tortuosity in retinopathy of prematurity using image analysis. *Med. Image Anal.* 6, 407–429 (2002)

5. Heras, E., De la Prieta, F., Julian, V., Rodríguez, S., Botti, V., Bajo, J., Corchado, J.M.: Agreement technologies and their use in cloud computing environments. *Progress in Artificial Intelligence* 1(4) (2012)
6. Hoover, A., Goldbaum, M.: Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Trans. Biomed. Eng.* 22, 951–958 (2003)
7. Kalvainen, H., Hirvonen, P., Xu, L., Oja, E.: Probabilistic and non-probabilistic Hough transforms. *Image Vision Comput.* 13, 239–252 (1995)
8. Lee, S., Wang, Y., Lee, E.: A computer algorithm for automated detection and quantification of microaneurysms and haemorrhages in color retinal images. In: SPIE Conference on Image Perception and Performance, vol. 3663, pp. 61–71 (1999)
9. Li, H., Chutatape, O.: Automated feature extraction in color retinal images by a model based approach. *IEEE Trans. Biomed. Eng.* 51, 246–254 (2004)
10. Patton, N., Aslam, T.M., MacGillivray, T., Deary, I.J., Dhillon, B., Eikelboom, R.H., Yugesan, K., Constable, I.J.: Retinal image analysis: Concepts, applications and potential. *Progress in Retinal and Eye Research* 25(1), 99–127 (2006)
11. Rodríguez, S., De Paz, J.F., Bajo, J., Corchado, J.M.: Applying CBR Systems to Micro-Array Data Classification. In: Corchado, J.M., De Paz, J.F., Rocha, M.P., Riverola, F.F. (eds.) *Proceedings of IWPACBB 2008*. ASC, pp. 102–111. Springer, Heidelberg (2010)
12. Sánchez, C., Hornero, R., López, M.I., Aboy, M., Poza, J., Abásolo, D.: A novel automatic image processing algorithm for detection of hard exudates based on retinal image analysis. *Medical Engineering and Physics* 30(1-3), 350–357 (2008)
13. Tamura, S., Okamoto, Y., Yanashima, K.: Zero-crossing interval correction in tracing eye-fundus blood vessels. *Pattern Recogn.* 21, 227–233 (1988)
14. Tanabe, Y., Kawasaki, R., Wang, J.J., Wong, T.Y., Mitchell, P., Daimon, M., et al.: Retinal arteriolar narrowing predicts 5-year risk of hypertension in Japanese people: the Funagata study. *Microcirculation* 17, 94–102 (2010)
15. Wong, T.Y., Duncan, B.B., Golden, S.H., Klein, R., Couper, D.J., Klein, B.E., et al.: Associations between the metabolic syndrome and retinal microvascular signs: the Atherosclerosis Risk In Communities study. *Invest Ophthalmol Vis. Sci.* 45, 2949–2954 (2004)

Real-Time Wrist Localization in Hand Silhouettes*

Tomasz Grzejszczak, Jakub Nalepa, and Michał Kawulok

Abstract. This paper is focused on wrist localization which is an important step in recognizing hand gestures. While there are many methods for detecting hand feature points as well as for estimating the hand pose, the majority of them assume that the palm region is given and ignore the wrist detection step. However, despite it is a required operation if the gesture recognition is supposed to be automatic, wrist localization has not been given much attention in the literature. Here, we propose a fast, yet effective method for wrist localization and we present the evaluation procedure based on our set of 899 hand images with ground-truth data. To the best of our knowledge, such a quantitative analysis of this problem has not been published so far.

1 Introduction

Hand gestures constitute a natural way of communication between humans in their everyday lives. Usually, they play a supportive role to the verbal messages, but in some cases these are the gestures which are used to convey the relevant information. Hand detection, tracking and gesture recognition have a number of practical applications including the real-time videoconferencing [11], telemedicine and medical visualization [15], telerobotics [1], interactive augmented reality [14], human-computer and human-robot interaction (HCI, HRI) interfaces [13, 17] and more [3, 12].

Faculty of Automatic Control, Electronics and Computer Science,
Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
e-mail: {tomasz.grzejszczak, jakub.nalepa}@polsl.pl,
michal.kawulok@polsl.pl

* This work has been supported by the Polish Ministry of Science and Higher Education under research grant no. IP2011 023071 from the Science Budget 2012–2013.

In this paper we propose a fast algorithm for automatic wrist localization from hand silhouettes, obtained after skin detection [9]. Contrary to the existing techniques, our method is independent from the hand orientation and does not impose any constraints on the sleeve length. The algorithm is based on the hand contour and width profile analysis which is performed using fast and simple operations. The experimental results prove that our method is effective and allows for real-time processing.

The paper is organized as follows. Existing wrist detection techniques are discussed in Section 2. The algorithm for the wrist points detection is described in detail in Section 3. The experimental results are given and discussed in Section 4. Section 5 concludes the work.

2 Related Literature

HCI interfaces based on hand gestures analysis have been extensively studied during the recent years in order to simplify human interaction with the surrounding environment. Two main gesture-based HCI categories can be distinguished: (1) hardware-incorporating [16] and (2) vision-based methods [4]. The former takes advantage of an additional hardware, e.g. magnetic sensors, gloves and markers, which is used for locating the hand, extracting its features and finally – recognizing the gesture. They offer high accuracy and robustness, however the excessive hardware cost, lack of naturalness and lowered comfort of use, make these methods virtually inapplicable in case of the solutions designed for ordinary customers. Contrary to that, the latter group of vision-based approaches relies exclusively on the acquired images to extract hand features and to classify the gestures. Although they are contact-free and comfortable for users, these methods are highly dependent on the environment, e.g. the lighting. Thus, the additional effort needs to be made to design precise and robust hand detection and analysis algorithms applicable in real-time systems.

Skin region segmentation and hand detection constitute the primary steps in the case of the vision-based methods. Correct and precise detection is crucial for subsequent pose estimation, features extraction and classification. In general, a human hand is composed of three parts: fingers, palm and wrist. There exist methods ignoring the wrist detection to simplify the processing, however, its position is a valuable source of information about the hand direction and position [6].

The width-based wrist location technique was used by Licsár and Szirányi [10]. The widths of the forearm are evaluated along the main arm direction. According to the anatomical structure information of the hand, the width starts increasing from the wrist (from the forearm to the palm). In this method it is necessary to find the main arm direction what requires time-consuming calculation of the image moments. Furthermore, the main arm direction cannot be determined correctly if the arm is bent in the elbow.

Dung and Mizukawa proposed a distance transform-based method [5] for determining the palm region, which was used in our earlier gesture recognition system [4]. The wrist line can be determined with a high accuracy here, but it is assumed that only a short part of the forearm is presented.

Hu et al. proposed a color-based approach [8]. In this technique, an image containing the hand is divided into 9 sub-blocks with the fist in the center. Due to a high value of correlation between the hand and wrist colors, the average value of color of the block with the wrist is the most similar to the central block. Using the circular projection of the hand area to the wrist detection has been recently proposed by Hasan and Mishra [7]. The main drawback of this method is the assumption that the palm center and the fingers area are known. Thus, the technique can be applied only for the wrist locating in the pre-processed images with determined positions of the hand parts.

3 Proposed Wrist Localization Algorithm

Our method is based on an observation that the wrist forms a local minimum in the width profile of the hand silhouette. The profile is computed along the longest chord that can be determined inside the hand region. Using this rule, illustrated in Fig. 1, the wrist position can be correctly located for most cases of hand silhouettes, independently from the hand orientation and size of the skin blob, which is an important advantage over existing techniques. The silhouettes were obtained using a skin detector developed earlier during our research works [9], and also for the sake of validation the ground-truth masks were used, which is explained in Section 4.

At first, the longest chord (segment PQ , where Q lies in the palm) is determined and the silhouette is rotated, so as the segment is positioned horizontally. After rotation, the width at every position of the rotated chord is obtained as a sum of skin pixels in each column of the rotated image. Profile analysis makes it possible to determine the wrist position, which is projected back to the original image afterwards and appears as the segment UV in Fig. 1. Central point of this segment is treated as the detected wrist position. This algorithm is outlined in Alg. 1 and its details are given later in this section.

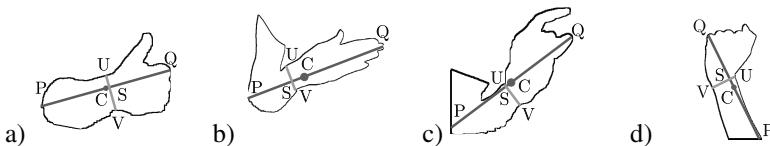


Fig. 1 Hand silhouettes with the longest chords (PQ) and wrist lines (UV) annotated

Algorithm 1 Automatic wrist points detection

- 1: Prepare input data: scale the image to a size of 300×300 pixels, calculate the contour;
 - 2: Determine the longest chord of the contour;
 - 3: Rotate the image by an angle of the chord's slope;
 - 4: Find local extrema, assuming local minimum to be the wrist point;
 - 5: Compute the final wrist point in the original image;
-

3.1 Hand Orientation

An input color image is subject to skin detection. Every detected skin blob is down-scaled to a maximum size of 300×300 pixels. This does not affect the effectiveness while it makes the further processing time independent from the input image size.

Hand orientation is determined based on the longest chord PQ which satisfies the convexity assumption, i.e. $PQ \subseteq \mathcal{M}$, where \mathcal{M} is the set of skin pixels (see Fig. 2 where $P''Q'' \not\subseteq \mathcal{M}$ and $P''Q \not\subseteq \mathcal{M}$). The longest chord can be determined using a brute-force approach, in which every chord PQ is verified. However, this results in an average processing time of 49 seconds per image. Thus, a randomized approach was applied which significantly reduces the search time (~ 550 times). The proposed algorithm for finding the longest chord is presented in Alg. 2. It is initiated with a randomly chosen chord defined by two points P and Q from the contour V (line 4). A new pair of points is chosen randomly from the contour and it is checked whether they form the longest and valid ($PQ \subseteq \mathcal{M}$) chord (line 5). If they do, then a new iteration is initiated (line 7). In every iteration, a maximal number of L random chords are found. If none of them updates the chord determined in the previous iteration, then the local search is performed (line 13). It consists in checking additional eight chords (lines 21-30), determined by the points that lie at a small distance Δ along the contour from the current end points (see Fig. 2). If the chord is updated, then the next iteration starts (line 14). If the local search fails to improve the result, then the search is stopped (line 16).

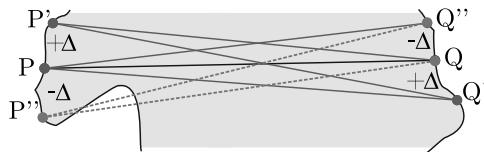


Fig. 2 Local optimization for improving the longest chord (PQ is changed to $P'Q'$)

Theoretically, the randomization involves a risk that the search may be very time consuming or that the determined chord is far from the optimal one. However, an extensive experimental study indicated that the chords were determined correctly and

that the search process never exceeded 11 iterations. In the worst case, the longest chord was found after 0.36 seconds, which is only 4.4 times slower than the average, and over 100 times faster than the brute force.

Algorithm 2 Longest chord search

INPUT: $\mathbf{V}, \mathcal{M}, \Delta, L$ \triangleright a vector of contour points (\mathbf{V}) and a set of skin pixels (\mathcal{M})

- 1: $d_{max} \leftarrow 0;$ \triangleright length of the longest chord
- 2: $stop \leftarrow \text{FALSE}; counter \leftarrow 0;$ \triangleright stop condition and chord counter
- 3: **while** $stop = \text{FALSE}$ **do**
- 4: $P = \text{random}(\mathbf{V}); Q = \text{random}(\mathbf{V});$ \triangleright get two random points from the contour
- 5: **if** $|PQ| > d_{max} \wedge PQ \subseteq \mathcal{M}$ **then** $\triangleright PQ$ becomes the longest chord
- 6: $d_{max} \leftarrow |PQ|;$
- 7: $counter \leftarrow 0;$ \triangleright starts a new iteration
- 8: $(P_{max}, Q_{max}) \leftarrow (P, Q);$
- 9: **else**
- 10: $counter \leftarrow counter + 1;$
- 11: **end if**
- 12: **if** $counter = L$ **then** \triangleright maximal number of draws in a single iteration
- 13: **if** LOCALSEARCH($P, Q, \mathbf{V}, \mathcal{M}$) = TRUE **then**
- 14: $counter \leftarrow 0;$ \triangleright the local search improved the result
- 15: **else**
- 16: $stop \leftarrow \text{TRUE};$ \triangleright stop condition is met
- 17: **end if**
- 18: **end if**
- 19: **end while** $\triangleright P_{max}Q_{max}$ is considered as the longest chord
- 20: **function** LOCALSEARCH($P, Q, \mathbf{V}, \mathcal{M}$)
- 21: $i \leftarrow \text{INDEX}(\mathbf{V}, P);$ \triangleright retrieves the index of the point P in the contour \mathbf{V}
- 22: $j \leftarrow \text{INDEX}(\mathbf{V}, Q);$ \triangleright retrieves the index of the point Q in the contour \mathbf{V}
- 23: $P' \leftarrow \mathbf{V}[i - \Delta]; P'' \leftarrow \mathbf{V}[i + \Delta]; Q' \leftarrow \mathbf{V}[j - \Delta]; Q'' \leftarrow \mathbf{V}[j + \Delta];$
- 24: $(X, Y) = \arg \max \{|XY| : X \in \{P, P', P''\} \wedge Y \in \{Q, Q', Q''\} \wedge XY \subseteq \mathcal{M}\};$
- 25: **if** $|PQ| < |XY|$ **then**
- 26: $(P, Q) \leftarrow (X, Y);$
- 27: **return** TRUE; \triangleright local search improved the result
- 28: **else**
- 29: **return** FALSE; \triangleright the initial segment PQ was locally the longest
- 30: **end if**
- 31: **end function**

3.2 Locating the Local Extrema

The longest chord PQ determines the direction, along which the widths of the silhouette are calculated. In order to simplify and speed up the computations, the image is rotated by an angle α , obtained from the slope of a line containing the segment PQ , around the center of rotation C (see Fig. 3). Thanks to this operation, the width profile can be computed by summing the mask pixels in each image column after rotation.

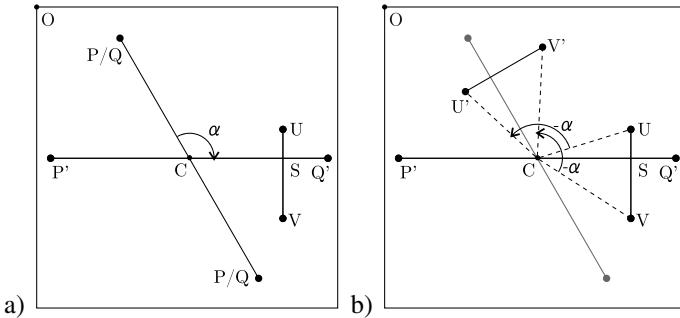


Fig. 3 Rotation of a silhouette

It is worth noting that the orientation of the segment PQ is unknown. Therefore, the palm area of a hand (i.e. point Q) can be located either on the left or on the right side in the rotated image. To address this problem, all contour points on both sides of the center of rotation C are counted after the rotation. It may be observed that the contour of the palm is usually longer than the opposite part. Given this assumption, we verify the image orientation and start the profile analysis from the palm region. There exist images, for which the number of contour points is smaller around the palm. In these cases, the proposed algorithm for profile analysis is slightly less effective, which is later explained in Section 4 and presented in Fig. 7.

The sum of skin pixels is computed for each position, i.e. for each column along PQ . The first local maximum found in the profile is located either in the palm area or in the vicinity of the elbow. Nextly, the first local minimum in the profile is interpreted as the wrist (S) in either case, although the results are better if the analysis is started from the palm. The minimum is calculated using the step contraction method and the directional minimum search [2]. In this approach, the step is geometrically reduced by the parameter k (line 7) until it is less than the defined accuracy ε (line 5). If the new value of the minimum is larger than the previous one, then the direction of the search is reversed (line 9). During the wrist search process, the local maximum only needs to be detected, rather than precisely located. Thus, its rough position is found by sampling the profile with a defined step δ , until the extremum is determined (lines 2-4). Finally, the location of U and V points is determined and mapped

to the input coordinate system as U' and V' . The most relevant steps of this procedure are presented in Fig. 4.

Algorithm 3 Wrist localization based on profile analysis

```

INPUT:  $S, \delta, \varepsilon, k, n_i$       ▷ a vector of the rotated mask profile  $S$  and the settings
1:  $x \leftarrow 0; n \leftarrow n_i;$           ▷  $x$  indicates the position in the profile
2: while  $S[x + \delta] > S[x]$  do
3:    $x \leftarrow x + \delta;$            ▷ find the local maximum
4: end while
5: while  $n > \varepsilon$  do           ▷ find the local minimum
6:   if  $S[x + n] < S[x]$  then
7:      $n \leftarrow n/k;$            ▷ contract the search step
8:   else
9:      $n \leftarrow -n/k;$            ▷ change the search direction
10:  end if
11:   $x \leftarrow x + n;$ 
12: end while
  
```

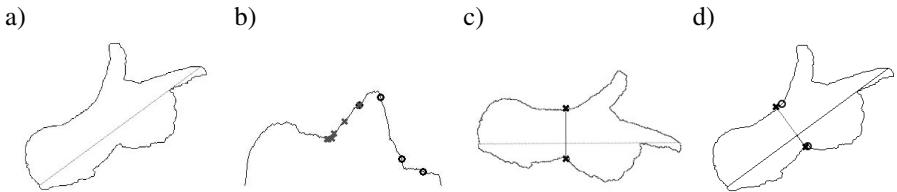


Fig. 4 Wrist localization steps: a) the longest chord, b) sampled minima (cross) and maxima (circle) of the profile, c) detected wrist points, d) detection result compared with the ground-truth data (circles)

4 Experimental Validation

The proposed method was validated using our database¹ containing 899 images with the ground-truth data [9]. The parameters were tuned experimentally to the following values: $L = 50$, $\Delta = 5$, $\delta = n_i = P_w/10$, where P_w is the profile width, $\varepsilon = 2$, $k = 1.5$. The algorithm was implemented in C++ and run on an Intel Core i5 2.3 GHz (4 GB RAM) computer.

In order to verify the performance of our approach, we detected the points U' , V' and W' for each image, where W' is the wrist point, and compared them with the ground-truth points U , V and W . The detection error e is defined as $e = |WW'|/|UV|$,

¹ The data set is available at <http://sun.aei.polsl.pl/~mkawulok/gestures>.

$e \leq E$, where E is the maximal detection error ($E \leq 1.0$). Possible wrist point positions are given in Fig. 5 (indicated as dashed circles for $E = 0.5$ and $E = 1.0$). It is worth noting that the position of the point W' is constrained by the positions of U' and V' laying on the contour, as it is the middle point of $U'V'$.

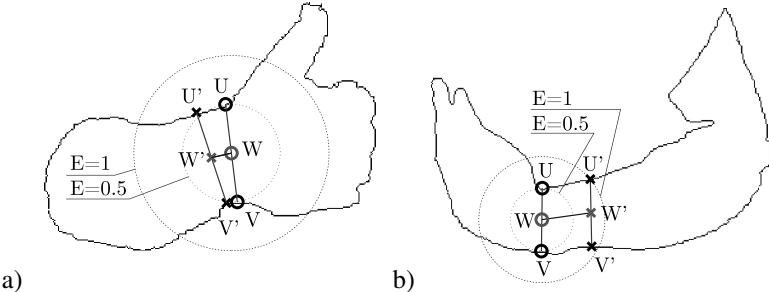


Fig. 5 Silhouette with the ground-truth (U, V, W) and detected (U', V', W') points and possible wrist point areas. The detection errors are a) $e = 0.2091$ and b) $e = 0.7887$.

The wrist in a given image is considered as properly detected if $e < E$. Thus, the accuracy of detection can be calculated either for the entire database (including images for which $e \geq E$) or only for those correctly detected. Processing times and errors obtained for $E = 0.5$ and $E = 1.0$ are given in Tab. 1. Here, the results presented for the proposed method (termed *Random*) were averaged over 11 tests run using the entire data set. We compared them with 1) a brute-force approach for finding the longest chord outlined earlier in Section 3.1 (*Brute Force*), and 2) the algorithm proposed by Licsár and Szirányi [10], based on the image moments calculation (*Moments*).

Table 1 Comparison of time and error statistics

		Random	Brute Force	Moments
Processing time	t_{avg}	0.0812 ± 0.0322	49.680 ± 27.276	0.0254 ± 0.0032
	t_{max}	0.359	176.111	0.042
Total error	e_{avg}	0.5940 ± 0.6791	0.5937 ± 0.6832	0.6016 ± 0.6563
$E = 1.0$	Number of $e > E$	131.7 (14.7%)	132 (14.7%)	111 (12.4%)
	$e_{avg} > E$	0.3005	0.2990	0.3439
$E = 0.5$	Number of $e > E$	322.3 (36.0%)	309 (34.4%)	370 (41.3%)
	$e_{avg} > E$	0.1547	0.1604	0.1487

An average t_{avg} and the worst t_{max} processing time of the proposed method is significantly reduced compared to the brute-force approach, without a significant loss of accuracy. Given a small value of the standard deviation, the maximal time t_{max} of our method can be considered as a gross error ($0.1 \leq t \leq 0.3$ was observed for approx. 0.05% cases). The processing time histogram for the *Random* algorithm

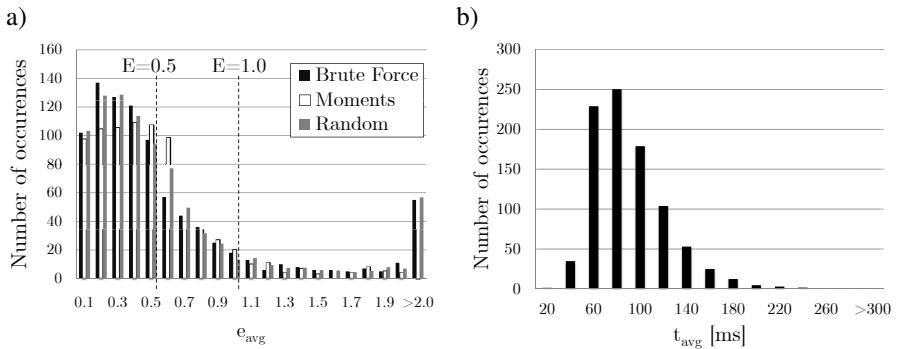


Fig. 6 Histograms of a) the average detection error e_{avg} for 11 tests and b) the average execution time t_{avg} of the *Random* algorithm for 11 tests

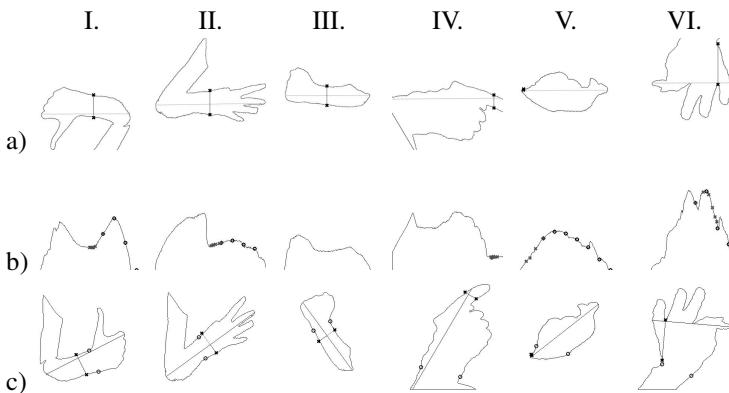


Fig. 7 Examples of wrist localization: a) image rotated based on the detected chord, b) width profile of (a), c) obtained result compared with the ground-truth location

is given in Fig 6b. The *Moments* algorithm turns out to be the fastest, but the detection accuracy is affected due to imprecisely determined rotation angle. This can be noticed for $E = 0.5$, where the average number of improperly detected wrist points grows rapidly from 12.4% for $E = 1.0$ to 41.3%. The error rates are similar in case of the random and brute-force approaches (Fig. 6a). It is worth noting that an average error value e_{avg} of the best results obtained using the *Random* algorithm equals 0.2 (with 1.5% and 7% of incorrect results for the maximal error $E = 1.0$ and $E = 0.5$, accordingly). Several examples of obtained outcome are presented in Fig. 7. The most important cases are covered here, namely: correct wrist localization in case of an incorrect detection of the palm orientation (I.), images with the entire hand visible (I.–III.), images presenting hands with long sleeves (IV.–V.), and incorrect outcomes (IV., VI.).

5 Conclusions and Future Work

In this paper we proposed a fast algorithm for wrist localization in hand silhouettes based on the profile analysis along the longest chord, determined using a randomized approach. Experimental results proved that the effectiveness of our method is competitive compared to an alternative existing algorithm, while the execution time allows for real-time processing, contrary to the brute-force technique.

Our ongoing research includes developing more advanced profile analysis techniques, which may help decrease the detection errors. Furthermore, the failures of the presented algorithm are very often observed for hands with a long sleeve, where the wrist point is located on the contour, and this problem is planned to be addressed as well.

References

1. Al-Mouhamed, M., Toker, O., Al-Harthy, A.: A 3-D vision-based man-machine interface for hand-controlled telerobot. *Ind. Electr.* 52, 306–319 (2005)
2. Bukowiecki, B., S.S., Galuszka, A.: Optimization of the geometrical features of the powered roof support unit kinematic chain due to the criterion of minimizing the weight. *Electr. Rev.* 3a, 131–138 (2012)
3. Collumeau, J.-F., Leconge, R., Emile, B., Laurent, H.: Hand-gesture recognition: Comparative study of global, semi-local and local approaches. In: Proc. IEEE ISPA, pp. 247–252 (2011)
4. Czupryna, M., Kawulok, M.: Real-time vision pointer interface. In: Proc. ELMAR, pp. 49–52 (2012)
5. Dung, L., Mizukawa, M.: Fast hand feature extraction based on connected component labeling, distance transform and hough transform. *J. of Rob. and Mech.* 21, 726–738 (2009)
6. Hasan, M.M., Mishra, P.K.: Hand gesture modeling and recognition using geometric features: A review. *Canadian J. on Im. Proc. and Comp. Vision* 3, 12–26 (2012)
7. Hasan, M.M., Mishra, P.K.: Novel algorithm for multi hand detection and geometric features extraction and recognition. *Int. J. of Sc. and Eng. Res.* 3, 1–12 (2012)
8. Hu, K., Canavan, S., Yin, L.: Hand pointing estimation for human computer interaction based on two orthogonal-views. *Patt. Rec.*, 3760–3763 (2010)
9. Kawulok, M.: Energy-based blob analysis for improving precision of skin segmentation. *Mult. Tools and App.* 49, 463–481 (2010)
10. Licsár, A., Szirányi, T.: Hand gesture recognition in camera-projector system. In: Int. Workshop on Human-Computer Int., pp. 81–91. Springer (2004)
11. MacLean, J., Pantofaru, C., Wood, L., Herpers, R., Derpanis, K., Topalovic, D., Tsotsos, J.: Fast hand gesture recognition for real-time teleconferencing applications. In: Proc. IEEE ICCV, pp. 133–140 (2001)
12. Mitra, S., Acharya, T.: Gesture recognition: A survey. *IEEE Trans. on Systems, Man and Cybernetics - Part C* 37, 311–324 (2007)
13. Sales Dias, M., Gibet, S., Wanderley, M.M., Bastos, R. (eds.): *Gesture-Based Human-Computer Int. and Sim.* Springer, Heidelberg (2009)

14. Shen, Y., Ong, S.-K., Nee, A.Y.C.: Vision-based hand interaction in augmented reality environment. *Int. J. Hum. Comput. Interaction* 27, 523–544 (2011)
15. Wachs, J., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M., Handler, J.: A real-time hand gesture interface for medical visualization applications. *App. of Soft Comp.* 36, 153–162 (2006)
16. Wei, F., Xiang, C., Wen-hui, W., Xu, Z., Ji-hai, Y., Lantz, V., Kong-qiao, W.: A method of hand gesture recognition based on multiple sensors. *Bioinf. and Biomed. Eng.*, 1–4 (2010)
17. Zhang, J., Zhao, M.: A vision-based gesture recognition system for human-robot interaction. In: Proc. IEEE ROBIO, pp. 2096–2101 (2009)

An s -layered Grade Decomposition of Images

Maria Grzegorek

Abstract. An s -layered model of the grade data analysis is used to image decomposition. The method is derived from grade correspondence analysis and involves grade outliers. Image pixels are described by selected variables. The resulting data table is ordered according to the maximal grade differentiation. The outlying measure from the main trend is calculated for each pixel. This measure defines pixels distances from the regularity. The data table ordered according to outlying measure is divided into more homogeneous subsets and subsets form subimages with more similar pixels (in grade outlier meaning).

1 Introduction

The Grade Correspondence Analysis (GCA) has recently been extensively applied to the data exploration. The basis of grade methods and models is described in [5]. Many simple examples illustrate variety of cases. Some practical tasks are also mentioned. (For example analysis of questionnaires concerning disabled people where employment is related to functional ability, skills, education, age and so on. Another example concerns high performance concrete.) An easy introduction and some interesting dependencies between EU countries can be found in [6]. Images also undergo GCA if only data table is formed involving any information from the image. Previous articles (for example [4] show how to form data table which rows are pixels and columns are variables describing some features of pixels. The number of variables has not occurred very large that profits in time shortening. Also variables construction can be simplified by avoiding parameters and thresholds. Variables depending upon values of certain simple features of neighboring pixels behave good enough.

Maria Grzegorek

The Institute of Computer Science PAS, Jana Kazimierza 5, 01-248 Warsaw, Poland

e-mail: mary@ipipan.waw.pl

The last required parameter, the number of clusters in grade correspondence cluster analysis is omitted in a version of grade-spatial algorithm ([3]). Search for significant groups of pixels is simultaneously performed in grade space (grade order of pixels) and in image space (adjacency of pixels in the image).

This text profits by s -layered grade inference framework [7]. In a typical grade data exploration objects are divided into two more or less even subsets, one fitting to main trend and another outlying from such tend, maybe fitting to a second trend. The partition is made with the use of statistics called *AvgDiffA*. This measure defines an average difference for each object from remaining objects in grade ordered data table. Objects in turn are ordered from the most to the least outlying and a partition point is selected from a graph of *AvgDiffA*. This time only objects with the smallest *AvgDiffA* values are separated and remaining objects undergo the procedure from the beginning. We get s layers of image pixels which form more regular clusters.

Arbeláez at al. [2] perform contour detection task in its local part in four channels, e.g. three CIE Lab colours and texture. In those channels every pixel is a center of a disc of some diameter. Distance of two histograms derived from both half-discs is a measure of gradient in direction orthogonal to diameter. Texture channel requires some preliminary calculation which involves convolution with Gaussian filters and attachment the number of relative texton. Then the gradient for texture channel is calculated in the same way. Multiscale measure is obtained by performing weighted sum of gradients over all discs diameters and all channels. The globalization feature is gained when sparse affinity matrix is formed involving maximum gradient along a path connecting every two pixels in the fixed distance. Eigenvectors for the smallest eigenvalues, which are images too, are convolved with Gaussian directional filters and summed up. The final weighted sum of multiscale and spectral component is evaluated with the aid of Berkley Segmentation Dataset (BSDS).

Antúnez at al. [1] divide the task into two stages. In a pre-segmentation stage combinatorial pyramid is build with established levels of combinatorial maps involving darts, ordered sets of darts belonging to vertices, sets of darts belonging to the same arc. The base level includes all pixels. Contraction or removal is based on merging patches which colour difference (region feature) is below a threshold. The perceptual grouping stage depends on removal boundaries between patches with strength below another threshold (boundary features). The results show high topological information maintenance. Evaluation involving developed in [2] values of precision and recall measures uses BSDS and human-created segmentation. Widynski and Mignotte [8] develop contour detector involving Bayesian model, particle tracking technique and edgelets (short fragments of contours) as a subject of distribution transfer. Authors use BSDS in both stages: learning phase and contour detection.

Section 2 introduces theoretical background. Section 3 presents results. Section 4 sums up performed task and outlines improvement of proposed processing.

2 Grade Correspondence Analysis and Grade Outliers

Grade analysis involves ideas of transforming multivariate dataset onto bivariate copula as well as ideas of regularity and strength of monotonic dependence in data tables. To take advantage of this concept the set of image pixels is transformed into a data table with size $m \times k$. Rows in the data table are m image pixels. Each pixel is attached by k values of variables. Pair (i, s) represents table cell containing value of variable s attached to pixel i .



Fig. 1 Test image "Lena" (left), four selected pixels (right)

Figure 1 shows one of test images and a magnified fragment of the test image with four pixels marked in white. Pixels characteristics are listed in Table 1 (from left to right on the image 1: theirs numbers of row/pixel i , coordinates x , y and variable values $m1 \text{div} m8$. Numerical characteristics of pixels are transformed: each value in a row is divided by the sum of that row. Obtained values are shown in Table 2. Next the accumulated values are calculated for every pixel/row. The current values are shown in Table 3.

Table 1 Variables values for four selected pixels

i	x	y	gr.m	gray	m1	m2	m3	m4	m5	m6	m7	m8
32377	146	104	20,10	35	35	37	38	39	40	40	42	47
32378	147	104	120,42	38	35	35	40	40	42	55	68	87
32381	150	104	138,31	65	50	56	65	77	86	87	90	112
32386	155	104	28,46	204	191	198	199	202	204	204	204	204

Pixel with coordinate $x = 146$ is compared to three other on graphs illustrated in Figure 3 (upper row of graphs). Graphs are unit squares with diagonals which end points are $(0,0)$ and $(1,1)$. Consecutive variable values for pixel 146 are on x axis and values for the second, third and fourth pixel on y axes on successive graphs. Resulting curve is so called a concentration curve. Concentration curve is a tool used for comparison of two univariate probability variables and here illustrates dissimilarity of two pixels. The more distant concentration curve is from the diagonal the bigger dissimilarity occurs between two pixels. Thus the concentration curve

approaching to diagonal means bigger similarity of distributions attached to pixels. On the diagonal distributions are identical. The diagonal is also baseline reference in the visualization of concentration index. The greater area is associated with a greater dissimilarity. The area between the concentration curve and the diagonal is a measure of dissimilarity between two pixels. On shown plots the concentration curve is partly above the diagonal, partly below the diagonal. Concentration index denoted ar is equal to double difference between these two areas.

Table 2 Normalized values of variables for selected pixels

x	y	gray	gr.m	m1	m2	m3	m4	m5	m6	m7	m8	
32377	146	104	0,0539	0,0938	0,0992	0,1018	0,1045	0,1072	0,1072	0,1126	0,1260	
32378	147	104	0,2149	0,0678	0,0625	0,0625	0,0714	0,0714	0,0749	0,0981	0,1213	0,1552
32381	150	104	0,1674	0,0787	0,0605	0,0678	0,0787	0,0932	0,1041	0,1053	0,1089	0,1355
32386	155	104	0,0155	0,1110	0,1039	0,1077	0,1082	0,1099	0,1110	0,1110	0,1110	0,1110

Table 3 Cumulated values of variables for selected pixels

x	y	gray	gr.m	m1	m2	m3	m4	m5	m6	m7	m8	
32377	146	104	0,0539	0,1477	0,2415	0,3407	0,4425	0,5470	0,6542	0,7615	0,8740	1
32378	147	104	0,2149	0,2827	0,3451	0,4076	0,4790	0,5503	0,6253	0,7234	0,8448	1
32381	150	104	0,1674	0,2460	0,3066	0,3743	0,4530	0,5462	0,6503	0,7555	0,8645	1
32386	155	104	0,0155	0,1264	0,2303	0,3380	0,4463	0,5562	0,6671	0,7781	0,8890	1

Segments of the concentration curve can be ordered from left to right based on their slopes values, from small to large. The obtained curves are maximal concentration curves now and connected maximal concentration indices are called ar_{max} . The maximal concentration curve is convex and is always below the diagonal. Such maximal curves for pairs of above the mentioned pixels are shown in Figure 3 (middle row).



Fig. 2 Test images: crossroad (left), NMR brain image (right)

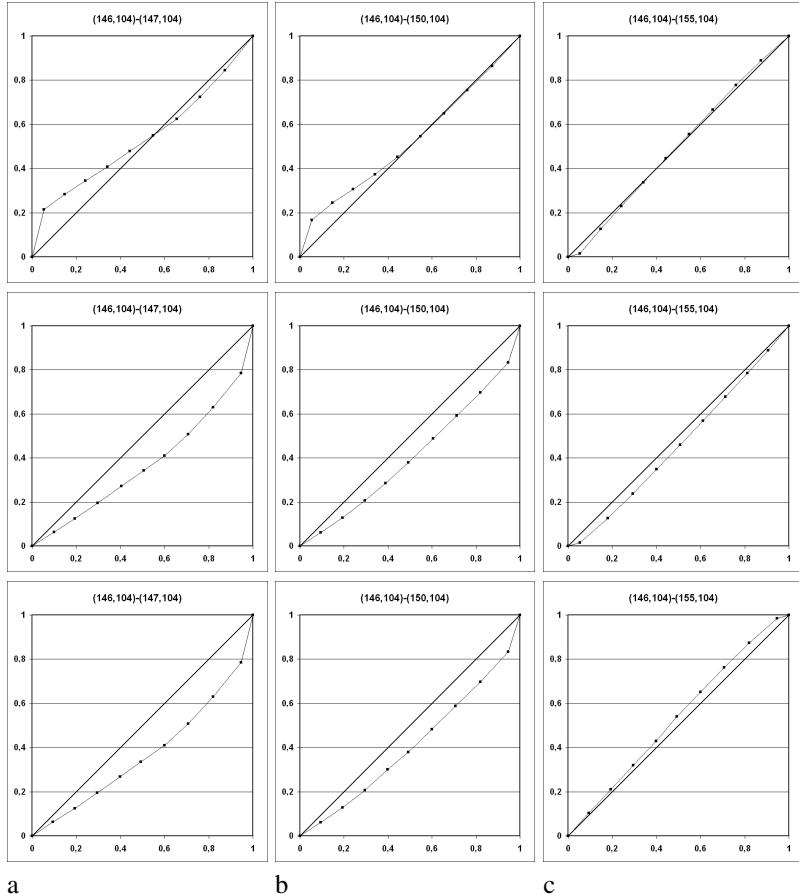


Fig. 3 Concentration curves for pairs of pixels from image 2: from left to right pairs a) (146,104):(147,104), b) (146,104):(150,104), c) (146,104):(155,104); top row - initial ordering of variables, middle row - maximal concentration curves, bottom row - after GCA

GCA allows to compare multivariate case of m image pixels attached by k values of variables. Value of variable in table cell (i,s) is divided by a sum of row i and is divided by a sum of column s . The data table gains a shape adequate to bivariate copula. Now the grade transform can be used to process the data table. Let p_{i+} be marginal sum for row i , p_{+s} is marginal sum for column s according to expressions:

$$p_{i+} = \sum_{s=1}^k p_{is} \quad (1)$$

$$p_{+s} = \sum_{i=1}^m p_{is} \quad (2)$$

Sums S_{i+} for row and S_{+s} for columns are:

$$S_{i+} = \sum_{j=1}^{i-1} p_{j+} + \frac{1}{2} p_{i+} \quad (3)$$

$$S_{+s} = \sum_{j=1}^{s-1} p_{+j} + \frac{1}{2} p_{+s} \quad (4)$$

Spearman ρ^* is a measure of differentiation of the data table:

$$\rho^* = 3 \sum_{i=1}^k \sum_{s=1}^m p_{is} (2S_{i+} - 1)(2S_{+s} - 1) \quad (5)$$

Spearmans ρ^* is frequently used as a measure of monotone dependence. It indicates how close is processed table to TP_2 - table which has features of Total Positivity of Order 2. To maximize ρ^* rows and columns are alternately sorted according to grade regression for rows $r_{row}(i)$ and for columns $r_{col}(s)$ [5]:

$$r_{row}(i) = \frac{\sum_{s=1}^k p(i,s) * S_{+s}}{p_{i+}} \quad (6)$$

$$r_{col}(s) = \frac{\sum_{i=1}^m p(i,s) * S_{i+}}{p_{+s}} \quad (7)$$

A new order of rows and columns with maximal ρ^* groups similar pixels while unlike pixels are found distant one from another in the data table. GCA is supplemented by clustering procedure which collects adjacent pixels in data table into clusters. Pixels in one cluster are similar and homogeneous and they are different if belongs to different clusters, especially if these clusters are distant one from another in the GCA ordered data table (extensive explanation of GCA mechanism can be found in [5], more compact in [6]). After GCA reordering regularity indices ar for pairs of rows increase but rarely reach maximal values. Increasing the global measure ρ^* by rows and columns reordering is a subtle balance between increasing and reduction of regularity indices for all pairs of pixels according to equation 5. Graphs of concentration curves after GCA procedure for selected pixels from Figure 1 are shown in Figure 3.

Afterwards, an idea is applied to the processed data table. Pixels are evaluated according to theirs outlying degree from the main trend revealed by procedure GCA. Such measure can be AvgDiffA statistics obtained according to equation 8. AvgDiffA is a mean distance between pixel and all remaining pixels in the data table reordered according to GCA.

$$AvgDiffA(i) = \sum_{s=1}^m \frac{(ar_{max}(i:s) - abs(ar(i:s)))}{(m-1)\sqrt{2}}, i = 1, 2, \dots, m \quad (8)$$

Rows are ordered from the most to the least outlying from regularity according to AvgDiffA. Graphs of AvgDiffA for test images from Figures 1 and 2 are shown in Figure 4, upper row.

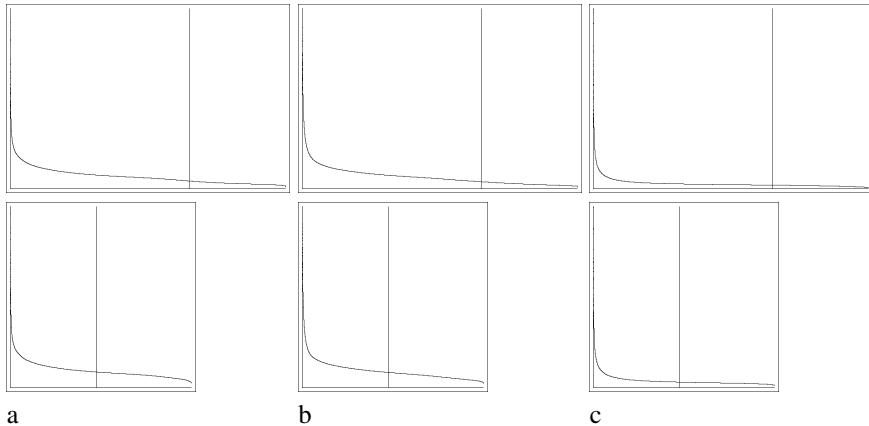


Fig. 4 AvgDiffA graphs: top - for all pixels, bottom - after omitting pixels attached to layer 1. On horizontal axes are objects (pixels), on vertical axes are values of AvgDiffA. The vertical lines show division points

Szczesny et al. [7] consider grade vertical composition of s contingency tables, when the number of columns is the same for each component. Particular cases/variables tables possessing properties which are "close" to the properties of two-way contingency tables can be vertically composed. In practical applications GCA could be applied to them which provide an overview of main trends. GCA can be followed by GCCA which forms a chosen number of ordered and possibly homogeneous clusters. Vertical decomposition of a table onto s components is a reversion of vertical composition. Model M_1 assumes decomposition onto two subsets: one fitting to the main trend revealed by GCA, second outlying from this trend. Decomposition is performed with the aid of AvgDiffA graph. The choice of a partition point is arbitrary and should take into consideration either similar number of pixels in both subsets or any noticeable point on the graph.

Another model called M_2 assumes s -layer decomposition. The layer is a regular but independent subpopulation detected in the original dataset. An iteration process is performed on data table: procedure GCA, AvgDiffA calculation, removal of most outlying object - until regularity index τ/τ_{max} gains value equal or close to 1 which means that remaining subset gains property close to TP_2 . In the case when the number of objects is large, a slightly different scheme is used. At the AvgDistA graph the partition point separates chosen number of pixels with the smallest AvgDiffA values. This subset forms first layer $M_2^{(1)}$. Remaining subset of pixels is reordered according to the procedure GCA, AvgDistA graph is composed and second layer

subset $M_2^{(2)}$ is formed. The process is continued s time until small set of the most outlying pixels remains.

3 Experiments and Results

The procedure involving s -layered decomposition is applied to the data table of pixels.

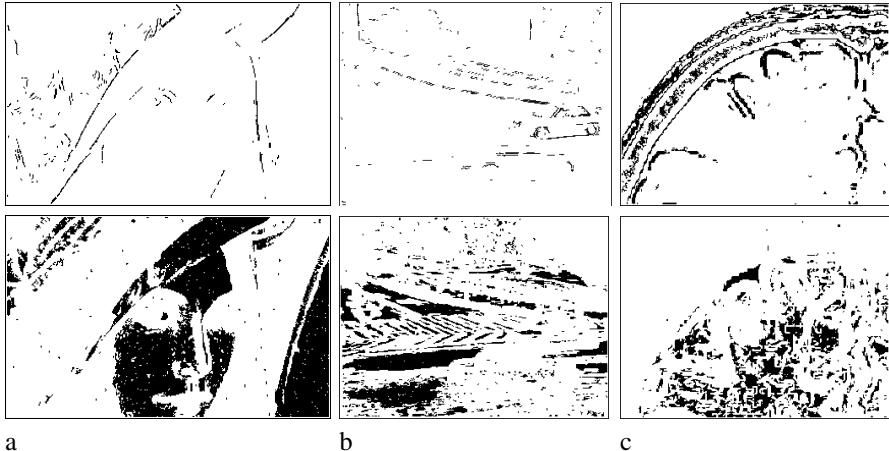


Fig. 5 Visualization of first subimages - top, and last subimages - bottom for test images

Gray test images and some gray images from Berkeley Segmentation Dataset were analyzed with a specialized application which makes possible processing of a huge data tables. There are ten variables: gray shade, gradient magnitude and eight gray values of neighboring pixels. These eight values are ordered in non-decreasing way. Pixels in data table are reordered according to GCA. AvgDiffA graph is calculated and 1/3 least outlying pixels are separated. Remaining pixels are GCA re-ordered and AvgDiffA values are determined again. These values for the reduced sets are shown on graphs in Figure 4 at the bottom row. Pixels from the subset ordered according to decreasing AfgDiffA values are divided into two subsets. So there are two layers ($s = 2$) and a set of outliers. Each subset is GCA reordered and three clusters are found. In each cluster pixels space dependencies are restored - pixels belonging to one cluster are displayed in one subimage. In each subimage regions are found on a eight-neighborhood principle. Some pixels do not remain connected. However, the number of such pixels is lower then in grade-space decomposition [3] or earlier grade image decomposition [4].

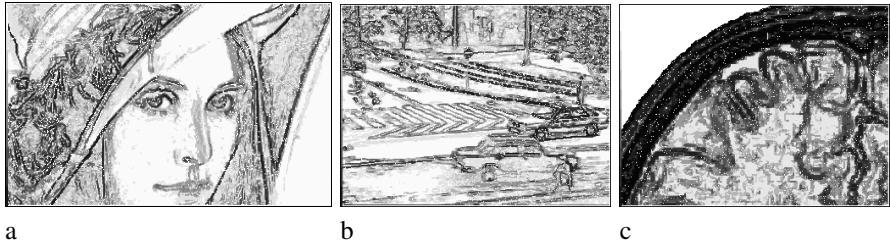


Fig. 6 Subimages visualization compacted in one image - contents of every subimage is drawn with separate colour/gray shade

Simple four characteristics are found for regions: mean gray value, standard deviation of gray, mean gradient magnitude and standard deviation of gradient magnitude. Region data table is reordered according to GCA and clusters are found. Regions belonging to one cluster are displayed in one subimage. Figure 5 presents some subimages: the first is like to a contour image and the last is segmentation image alike.



Fig. 7 Upper row: BSDS image, clusters visualization on one image; bottom row: visualization of the first and the last clusters

Figure 6 presents the results of decomposition displayed in one image. Pixels from one subset are displayed in one gray shade. Two pixels wide legend is put on left side of images. Figure 7 shows image of BSDS (top left), image visualizing all clusters (top right) and images of the first and the last cluster. Comparison between obtained results and human-performed segmentation/boundary detection is difficult. Achieved task results in disjoint, possibly meaningful, subsets of pixels whereas BSDS includes images either containing a few smooth segments or theirs smooth edges. Therefore, the obtained results of F-measure near 0.26div0.33 when only an

image of the first cluster is compared with human-created, seems promising enough to develop the following regions extraction and merging into clusters.

4 Conclusions

The method involving grade correspondence cluster analysis with s -layered decomposition is applied to gray images. Images are transformed into the data tables with m rows/pixels and $k = 10$ variables. Data table is reordered according to GCA procedure. Dataset is divided into subsets by successive subsets separation containing pixels with a low value of outlying measure. Each subset is processed by GCCA procedure which puts pixels belonging to it in grade order and divides subsets into smaller subsets which are more homogenous and more uniform. Pixels in subsets recover their space place and are displayed in separate subimages.

The subimages seem to be too rough units. Variables used to merge regions of pixels in bigger superregions are too little productive. There is a need to search another set of variables for regions. Another stage of method, AvgDiffA calculation is a significantly time consuming process. It demands either to accelerate algorithm or to apply more memory.

References

1. Antúnez, E., Marfil, R., Bandera, A.: Combining boundary and region features inside the combinatorial pyramid for topology-preserving perceptual image segmentation. *Pat. Rec. Let.* 33(16), 2245–2253 (2012)
2. Arbeláez, P., Maire, M., Fowlkes, C., Malik, J.: Contour Detection and Hierarchical Image Segmentation. *IEEE Trans. PAMI* 33(5), 898–916 (2011)
3. Grzegorek, M.: Grade-Spatial Procedure in Grade Decomposition of Medical Images. *Journal of Medical Informatics & Technologies* 17, 203–209 (2011)
4. Grzegorek, M.: Variable complexity reduction in image grade decomposition. In: Burdak, R., Kurzyński, M., Woźniak, M., Żołnierk, A. (eds.) *Computer Recognition Systems 4. AISC*, vol. 95, pp. 239–247. Springer, Heidelberg (2011)
5. Kowalczyk, T., Pleszczyńska, E., Ruland, F.: Grade Models and Methods for Data Analysis. In: Kowalczyk, T., Pleszczyńska, E., Ruland, F. (eds.) *With Applications for the Analysis of Data Populations. STUDFUZZ*, vol. 151, pp. 63–70. Springer, Heidelberg (2004)
6. Lenkiewicz, S.: Grade Data Analysis - The Concept and an Instance of Application. *Współczesne Problemy Zarządzania*, No. 1 (2012) (in Polish)
7. Szczesny, W., Kowalczyk, T., Wolińska-Welcz, A., Wiech, M., Dunicz-Sokolowska, A., Grabowska, G., Pleszczyńska, E.: Recent developments concerning models and methods of Grade Data Analysis (2012) (in print)
8. Widynski, N., Mignotte, M.: A Particle Filter Framework for Contour Detection. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012, Part I. LNCS*, vol. 7572, pp. 780–793. Springer, Heidelberg (2012)

System-Level Hardware Implementation of Simplified Low-Level Color Image Descriptor

Paweł Forczmański and Piotr Dziurzański

Abstract. The paper addresses the problem of Content-Based Image Retrieval implemented in hardware. The presented approach consists of a simplified low-level descriptor for color images and similarity measure that are implemented in a system-level C-based hardware description language ImpulseC. The floating to fixed-point data conversion has been applied and also some modifications, improving execution of loop instances parallelly has been described. The descriptor has been implemented as a core in a NoC-based MPSoC. The results of the implementation in an FPGA has been provided.

1 Introduction

An important area of modern computing is the processing and analysis of multimedia data: images, video and audio streams. Due to the growing size of multimedia streams used in modern computer systems, processing with conventional programming methods encounters the barrier that limits the performance of such systems. Hence, there is a strong need for more efficient computing models, which are characterized by a high degree of potential parallelism and, thus, may be achieved directly in dedicated hardware structures. Most of the proposed hardware and software solutions, aimed at multimedia processing, relatively rarely use the latest design of hardware systems, such as multi-core system-on-chip (MPSoC), network-on-chip (NoC) and transaction-level modeling (TLM).

Currently one of the most important yet difficult problems related to the processing of multimedia data is a recognition (in particular classification), which is used in many areas of modern knowledge-based economy, such as biometric security

Paweł Forczmański · Piotr Dziurzański
West Pomeranian University of Technology, Szczecin,
Faculty of Computer Science and Information Technology,
Żołnierska Str. 52, 71–210 Szczecin, Poland
e-mail: {pforczmanski, pdziurzanski}@wi.zut.edu.pl

systems, multimedia systems, indexing and retrieval, as well as human-computer communication and robotics.

The most important tasks in a typical multimedia data recognition are: preparation of a representative database and classification process. Construction of the database is a time-consuming process which involves creating representations of all possible reference objects. In the process of classification an object which is a subject to recognition undergoes several stages of processing: feature extraction for achieving the objective description, the dimensionality reduction of a set of features to create their compact representation and the final identification by assigning it to the appropriate class in the database [6]. A significant step towards a fully automated indexing systems has been made with the advent of MPEG-7 standard, which defines the various indexes (called descriptors) for multimedia data.

In the paper we present an approach to implementation of high-performance MP-SoC that will make it possible to improve the quality of multimedia data classification by increasing the number of criteria for the classification without increasing the computation time (due to the parallelization process).

Typically, the high accuracy of recognition is associated with a significant increase of system response time. The specificity of large collections of multimedia data (population of several thousand or more objects) is the fact that there is a high probability that the use of a single index (descriptor) does not guarantee a correct result. Therefore it is important to connect descriptors to enhance the effectiveness of the process. Such strategy is extremely sensitive to the structure of hardware implementation. In the case of software implementation of such algorithms, there are problems arising from the sequential nature of processing, in particular the time-consuming process of creating the feature space used in the identification process. On the other hand, in order to implement the above algorithms in hardware, it is necessary to extract data dependencies between various stages of processing and to determine memory access schedule.

The general objective of works described in this paper is to develop a hardware implementation of selected descriptors and classifiers used in the multimedia data recognition. Some preliminary works were described in [7]. Here we focus on visual descriptors related to color representation of still images as it seems to be one of the most important problems nowadays [12].

2 Existing and the Proposed Approach

Each type of multimedia object representation has its advantages and disadvantages, and the choice is not always obvious. It depends on many factors such as: field of application, user requirements, conditions during acquisition process. Most of the works are focused on the visual descriptors related to shape and color, because they guarantee relatively high performance [2], [11].

An important feature of digital images, often taken into consideration, is the color. This is due to the construction of human visual system (HVS). In good lighting conditions, one draws special attention, firstly, to the brightness and color, and then to shape, movement, texture and other features. So far, many descriptors have been proposed, most of them based on the histograms (distributions) of colors in different color spaces. Today, in the era of the dissemination of the MPEG-7 [2], the greatest expectations are associated with descriptors combining color information and its distribution - SCD (Scalable Color), DCD (Dominant Color), CLD (Color Layout).

In the literature, the problem of hardware implementation of multimedia data recognition procedures is present in two sub-areas: feature extraction and actual classification. Due to the relative ease of implementation, many of the currently published works relate to the hardware implementation of selected MPEG-7 descriptors, e.g. [13] and [14]. The second trend concerns the implementation of the classifiers, which is more complex and less represented.

On the other hand, publications that describe the simultaneous implementation of both stages of processing: extraction and classification are rare, what is caused by a large individual complexity of each of these problems and the limitations of hardware structures.

The analysis of the problem shows, that for numerous applications it seems reasonable to compute a few descriptors from the same data [9]. The codes computing such descriptors can be then implemented in separate hardware cores of a multi-core System on Chip and then computed concurrently. These cores are often connected with Network on Chip (NoC), as the NoC architecture offers high throughput and favorable communication, as well as effective flow control mechanisms in order to avoid problems in access to resources [1]. In the case of a system where cores compute independent descriptors, the possibility of multicast routing, i.e., sending a stream of the same data parallelly to numerous cores, is of particular importance [3].

Since the multi-core SoC systems include a collection of blocks, they typically use a regular architecture consisting of cores implementing algorithm specified in a high-level language, and the router, which is typically connected to the adjacent nodes. In our work, we use a typical mesh-based architecture with wormhole switching extended with a path-level multicast routing, as described in [4]. In the wormhole switching each packet is divided into smaller packets, flits (flow control units) [5]. The whole flit is transmitted between adjacent routers in a single clock cycle, or a single request-acknowledgement cycle in case of an asynchronous execution.

The mentioned above architecture imposes an orthogonality between computation and communication in a SoC circuit. The only requirement for a NoC core is to follow the communication protocol during the communication with the router connected to this core. This communication scheme is rather straightforward and usually implements the simple request-acknowledgement-based asynchronous communication. In the described system we also follow this technique, as described, e.g., in [3].

3 Image Matching Algorithm

The algorithm of image matching is based on Dominant Color Descriptor being a part of MPEG-7 Standard [11] and an intersection of histograms. It has been simplified for a hardware implementation. The following section describes the details of creating color-based image descriptor and the method devoted to comparing pairs of descriptors.

Color descriptor is calculated for an input image X being represented as a three dimensional structure of size $M \times N \times K$ in the RGB color space ($K = 3$). In the first step we perform a conversion from RGB to HSV representation, in order to reduce between-channel correlation. Here, the components of HSV fall within the following ranges: $Hue \in [0, 359]$, $Saturation \in [0, 100]$, and $Value \in [0, 100]$. Then we perform color quantization, in order to reduce the number of colors and create a histogram with 72 bins. The following thresholds are used $\forall m \in M, \forall n \in N$:

for *Hue* component ($k = 1$):

$$X^{(q)}(m, n, k) = \begin{cases} 0 & \text{for } X(m, n, k) < 20 \text{ and } X(m, n, k) \geq 316, \\ 1 & \text{for } 20 \leq X(m, n, k) < 40, \\ 2 & \text{for } 40 \leq X(m, n, k) < 75, \\ 3 & \text{for } 75 \leq X(m, n, k) < 155, \\ 4 & \text{for } 155 \leq X(m, n, k) < 190, \\ 5 & \text{for } 190 \leq X(m, n, k) < 270, \\ 6 & \text{for } 270 \leq X(m, n, k) < 295, \\ 7 & \text{for } 295 \leq X(m, n, k) < 316, \end{cases} \quad (1)$$

for *Saturation* and *Value* components ($k = 2$ and $k = 3$):

$$X^{(q)}(m, n, k) = \begin{cases} 0 & \text{for } 0 \leq X(m, n, k) \leq 20, \\ 1 & \text{for } 20 < X(m, n, k) \leq 70, \\ 2 & \text{for } 70 < X(m, n, k) \leq 100, \end{cases} \quad (2)$$

Next, a three-dimensional histogram $H^{(3D)}$ of size $H \times S \times V$ is built for $X^{(q)}$, where $H = 8$, $S = 3$ and $V = 3$, and later transformed into a one-dimensional vector for the sake of simplicity:

$$H(9h + 3s + v) = H^{(3D)}(h, s, v), \quad (3)$$

where h, s, v represent quantization values and $h = 0, \dots, H$, $s = 0, \dots, S$ and $v = 0, \dots, V$. Finally, the values in H are sorted in the decreasing order and the first 8 bins are left unchanged, while the other are zeroed. Next, the values in H are normalized to one by dividing them by a sum of all values. The comparison of two descriptors H_A and H_B belonging to analyzed images A and B is performed by a simple histogram intersection:

$$D = \sum_{i=1}^{72} \min(H_A(i), H_B(i)). \quad (4)$$

If the value of D is close to one, it means the images are similar, while D tending to zero suggests different images.

The above approach gives a possibility to compare images regarding their geometrical transformations (scaling and rotation). It is also invariant to some extent to cropping and noising.

Since the above algorithm uses several stages that are independent (i.e. color quantization and histogram calculation), it can be parallelized in order to create a rather effective hardware implementation.

4 Hardware Implementation

Since the code for image matching algorithm described in the previous section has been written in C language, it seemed favorable to choose one of the existing C-based hardware description language to implement the hardware-targeted counterpart of the source code. Recently, one of the most popular languages of this kind is ImpulseC [10]. It is an extension of ANSI C with new data types, aimed at hardware synthesis, new functions and directives for steering the hardware implementation.

The code is executed in so-called processes, which communicate with other processes using streams, signals, shared memory and semaphores. The process communication by means of streams is treated as a primary technique, and it is particularly suitable for multi-core SoCs, where each SoC core can be viewed as a single process. As the processes are to be realized in hardware, they may benefit from various ImpulseC optimization techniques, such as loop unrolling or pipelining. The first of these techniques are quite important in our system, as numerous computations in data-dominated algorithms are independent each other, such as RGB to HSV conversion, where the values for each pixel can be computed in parallel. The only task of the designer is then to find loops whose iterations are data-independent, and put an appropriate ImpulseC directive to compute them concurrently. However, there is usually no enough resources for generating hardware for each iteration, thus some kind of clustering is necessary. This trade-off between the computation time and target chip area can be established during a series of experiments. This is also the path followed by the authors.

Another problem with hardware realizations is to implement floating-point computations. While contemporary reconfigurable device vendors offer mathematic coprocessor cores, which are also supported by ImpulseC, their hardware realizations occupy quite large portion of the target device and do not easily follow the NoC principles (there is no simple possibility of transferring data between a core which needs the floating point operations and a mathematic coprocessor using NoC communication scheme, and it is difficult due to the target chip area limitation to implemented a separate coprocessor for each processing core). Thus we also assumed the conversion between floating-point and fixed-point code.

Despite ImpulseC offers fixed-point data types and mathematic operations on them, we preferred to determine the fixed-point type parameters using the SystemC

standard [8], another C-based hardware description language, aimed more at simulation than synthesis (in contrast to ImpulseC). SystemC offers more sophisticated fixed-point support, where both parts of a number separated with radix point is practically of an arbitrary length. Moreover, both saturation and quantization can be performed in various manner. There also exists some packages, such as CoCentric System Studio from Synopsys, that can automatize to some extent the process of finding the most suitable fixed points based on the simulation of the input data and minimizing the errors between floating and fixed point implementation. We conducted a series of experiment aimed at error-minimization using a simulation of four typical benchmark images. Then, for these input data, we determined the mean square error between floating and fixed point computations and thus determined the most favorable position of the radix point. After obtaining these parameters, we transferred it into our ImpulseC implementation.

After some simple code modifications, aiming at improving its hardware realization, we obtained our final code to be implemented in hardware. These modification included division of the code into coarse-grain partitions to be implemented in parallel. Since such transformation is always a trade-off between computational time and resource utilization, we analyzed the impact of each module onto the final realization in terms of particular functional blocks, such as adders, multipliers etc., estimated DSP blocks (present in our target FPGA chip) and number of computational stages. The results of this analysis are presented in Table 1. From this table it follows that two modules, RGB to HSV conversion and Histogram calculation require the largest number of resources. As both these modules are to be executed for two pictures, there is a possibility of implementing them once (then their functionalities are executed sequentially for the first and the second image; the number of required computational stages increases, but the resource need is significantly lower) or twice (two pictures are converted and both histograms are generated in parallel). Histogram intersection, Histogram calculation, Quicksort and Normalization, despite their lower impact on the final results, can be implemented as one or two instances. Only the last functional module, Histogram intersection, is to be generated as a single instance.

To estimate the impact of these modifications into the target chip parameters, we used Stage Master Explorer tool from the ImpulseC CoDeveloper package. This tool computes two parameters, Rate and Max Unit Delay (MUD), which approximates the performance of future hardware implementation. It is worth stressing that these parameters are computed instantly, in contrast with long-lasting hardware implementation. Thanks to this property, a user is capable of performing a more sophisticated design exploration, analyzing various modifications of the code and their impact on the target hardware. There is also possibility of observing the dataflow graphs of the code implementation, as well as a few types of pipeline graphs. Since we aim at increasing the performance, we applied maximal level of possible coarse-grain parallelism, i.e., to generate multiple instances of each suitable module. After fine-tuning we got 225 stages and the MUD was equal to 128 units.

At the last stage, we used Xilinx ISE to perform an implementation of the core in Virtex5 FPGA device (XC5VSX50T, Virtex 5 ML506 Evaluation Platform) and got the following device utilization: 5610 slice registers, 12024 slice LUTs, 4398 LUT-FF pairs. It means that about 37 percent of the device has been used, leaving more than 60% for the router and the remaining cores. Thus, in order to implement more than two cores in a single chip, an FPGA chip of larger capacity may be necessary. However, a number of interesting research regarding Network on Chip property and hardware realization performance can be also performed with a two-core SoC.

Table 1 Resources utilization and stages requirement for particular modules

Resource	Total	RGB to HSV	Histogram Calculation	Quicksort	Normalization	Histogram Intersection
7-bit Adders	4	0	2	0	0	0
10-bit Adders	4	0	0	2	0	0
14-bit Adders	4	2	0	0	0	0
32-bit Adders	76	19	5	9	3	2
16-bit Multipliers	4	0	2	0	0	0
32-bit Multipliers	20	10	0	0	0	0
32-bit Dividers	24	10	1	0	1	0
2-bit Comparators	2	0	0	0	0	0
32-bit Comparators	146	24	30	15	2	2
Estimated DSPs	84	40	2	0	0	0
Number of Stages	429	76	67	49	12	8

5 Summary

In the paper we presented the problem of hardware-based Content-Based Image Retrieval. The proposed approach consists of a simplified low-level descriptor for color images, which employs dominant colors histogram and the similarity measure involving histogram intersection. Such methods yield a hardware implementation, that can be easily parallelized. After intensive parallelization, both at the fine- and coarse-grain levels, we still managed to fit in a rather small-sized FPGA device, leaving some area to a necessary NoC router and at least one more core. Increasing the number of sequential computation, we could even manage to fit a slightly larger number of cores.

Acknowledgements. This work was supported by The National Science Centre (NCN) within the research project N N516 475540.

References

1. Benini, L., de Micheli, G.: Networks on Chips: A New SoC Paradigm. *IEEE Computer* 35(1), 569–571 (2002)
2. Bober, M.: MPEG-7 Visual Shape Descriptors. *IEEE Transactions on Circuits and Systems for Video Technology* 11(6), 716–719 (2001)
3. Chojnacki, B., Maka, T., Dziurzanski, P.: Virtual path implementation of multi-stream routing in network on chip. In: Malyshkin, V. (ed.) *PaCT 2011. LNCS*, vol. 6873, pp. 431–436. Springer, Heidelberg (2011)
4. Czarnecki, L., Dziurzański, P.: Adaptive tree-based multicast routing in network on chip architecture. *Electronics - Constructions, Technologies, Applications* (4) (2010)
5. Dally, W.J., Towles, B.: Route packets, not wires: on-chip interconnection networks. In: *Proceedings of the 38th Conference on Design Automation*, Las Vegas, Nevada, United States, pp. 684–689 (2001)
6. Forczmański, P., Frejlichowski, D.: Strategies of shape and color fusions for content based image retrieval. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierk, A. (eds.) *Computer Recognition Systems 2. ASC*, vol. 45, pp. 3–10. Springer, Heidelberg (2007)
7. Frejlichowski, D.: Analysis of possible system-level hardware implementation of selected shape description algorithms. *Journal of Theoretical and Applied Computer Science* 6(4), 51–58 (2012)
8. IEEE, IEEE 1666 Standard for SystemC Language Reference Manual (2006)
9. Kuncheva, L.I.: Combining classifiers: Soft computing solutions. *Pattern Recognition: From Classical to Modern Approaches*, pp. 427–452. World Scientific Publishing Co., Singapore (2001)
10. Kalogeridou, G., Voros, N.S., Masselos, K.: System Level Design of Complex Hardware Applications Using Impulse. In: *Proceedings of the 2010 IEEE Annual Symposium on VLSI, ISVLSI 2010* (2010)
11. Manjunath, B.S., Ohm, J.-R., Vasudevan, V.V., Yamada, A.: Color and Texture Descriptors. *IEEE Transactions on Circuits and Systems for Video Technology* 11(6), 703–715 (2001)
12. Marchewka, A., Kozik, R.: Computer vision method for detecting adult-oriented content in images. In: Choraś, R.S. (ed.) *Image Processing and Communications Challenges 4. AISC*, vol. 184, pp. 19–24. Springer, Heidelberg (2013)
13. Sniatala, P., Kapela, R., Rudnicki, R., Rybarczyk, A.: Efficient hardware architectures of selected mpeg-7 color descriptors. In: *15th European Signal Processing Conference EUSIPCO 2007*, Poznan (2007)
14. Xing, B., Fu, P., Sun, Z., Liu, Y., Zhao, J., Chen, M., Li, X.: Hardware Design for Mpeg-7 Compact Color Descriptor Based on Sub-Block. In: *8th International Conference on Signal Processing*, Beijing (2006)

Reconstruction of Head Surface Model from Single Scan

Krzysztof Skabek and Dawid Łapczynski

Abstract. The methods for reconstruction of incomplete mesh scans of faces were presented. The aim is filling the mesh surface with mesh processing methods: surface and volumetric. The problem of fitting ellipsoid surface to a single face scan was resolved. The resultant application performs semi-automatic conversion of single scan to the closed mesh model. The comparison of the ellipsoid-filling method to another method of surface reconstruction using multiple scans was conducted.

1 Data Acquisition

The 3dMD scanner¹ was used to obtain the mesh data and textures of faces. 3dMDFace scanning system is mainly used in medicine. The device consists of two modules and three cameras are mounted in each module (fig. 1a). This makes the transport, installation and calibration of the system quick and easy. The most suitable application for the system is obtaining the surface models of human faces. It is often used in the planning the facial surgery. The resulting model is the 3D triangular surface of the face from ear to ear and the RGB texture (fig. 1b). The scanning range is approximately 150 cm and the average objects have about 30 cm diameter. The scanning process itself is rapid, comfortable and does not require too much commitment.

3dMD systems work on the basis of the triangulation method of the structural light and is characterized by high speed of operation. Each system can measure the object within 1.5 milliseconds. The average measuring distance errors are of tenths

Krzysztof Skabek

Institute of Theoretical and Applied Informatics PAS
e-mail: kskabek@iitis.pl

Krzysztof Skabek · Dawid Łapczynski
Cracow University of Technology

¹ <http://www.3dmd.com/category/3dmd-systems/3d-systems/>



Fig. 1 3dMDFace scanning system: a) measurement device, b) single mesh from scanner, c) reconstructed mesh model of head

of millimeter relative to the reference model. Moreover, the structured light used in the scanner makes this device neutral for the eyes. These advantages make the system also useful for medical purposes.

A special type of 3D scanning, which is mainly considered in this work is face scanning. In order to obtain the accurate scan quality a special attention should be paid to preserve the mapping quality of parts of the face such as the nose, ears, the chin, etc. There are other devices with the ability to scan the model from all sides, which gives the full representation of human head without the necessity of tiresome model reconstruction which introduces inaccuracy in the mapping surfaces. However, such devices are expensive and rare. In our case, it is important to provide the solution for full head reconstruction (fig. 1c) using parital surface data coming also from cheaper scanning devices such as simple universal scanners or Kinect. Less important is high accuracy of the measurement expressed in distance units – nowadays almost all scanners have the accuracy to tenths of a millimeter, which is sufficient perform to face scanning.

A couple of popular scanners were compared in the paper [4]. According to the article the scanner Konica-Minolta offers the highest quality. The second in rank is 3dMD system offers a bit worse representation of the model, but acquisition time (less than 1 second to complete the scan) and quality of texture is better in this case. These parameters are very important for scanning faces. Also the use of a larger number of cameras 3dMD system gives an advantage, since it allows a wider range of accurate representation, for example, will provide better ears and optionally the back side of the head.

2 Reconstruction Techniques

Mesh reconstruction, also known as model repair, is the process of removing artifacts from a geometric model in order to obtain a suitable model output for further processing by applications requiring a guarantee of quality for their input. There

is no single algorithm providing the solution for all possible requirements of mesh reconstruction.

Artifacts may arise during triangulation od scan points or even at converting NURB surface on a triangular grid, which necessary to perform the analysis of the shape of the model and its optimization. Algorithms of mesh reconstruction can be oriented to the surface (surface-oriented algorithms) or algorithms operatig on volume data (volumetric repair algorithms).

Artifacts which may be taken into account at reconstruction are as follow [1]: large scale overlaps, holes and isles, topological noise, overlaps and gaps, edge intersection of faces, vertices common for multiple faces, inconsistant orientations of adjacent faces, etc.

2.1 *Surface-Oriented Methods*

Algorithms categorized as surface-oriented work directly on the input mesh trying to remove the artifacts with explicit modification of surface topology. The methods used here are simple operations, such as: attraction of edges, division of the mesh grid, sewing parts of the mesh. These algorithms are simple concepts and easy to implement. Algorithms are fast, and the level of occupied memory is low. This approach, however, does not guarantee the desired quality of the output.

Consistant orientation of normals. The important requirement for surface-oriented algorithms is the consistant normal orientation. It can also improve the quality of volumetric algorithms. Typically, the normal orientation is conducted in accordance with the minimum spanning tree between adjacent slices of the mesh.

Surface-based hole filling. The algorithm for surface-based hole filling was proposed by Peter Liepa in 2003 [8]. Its aim is to create a simple triangular mesh that fills polygon $(p_0, p_1, \dots, p_{n-1})$ which is a hole in the mesh. The main steps in this method are: identification of the hole; triangulation of the hole; subdivision of the filling mesh; profiling of the mesh according to the border faces of the hole.

Conversion to the manifold. Another important issue is the conversion of mesh models to the manifold. Manifold can be defined as a topological space where every point is homeomorphic to its surrounding in Euclidean space. It can be achieved by removing individual vertices and edges, as described in [5]. The algorithm operates directly on the edges of the model. In the first stage the edges and vertices are classified regarding the number of the adjacent faces. Then the overlapping edges are separated and next the topologically consistant pairs of edges are connected.

Filling narrow holes. We used some ideas to solve the problem. The first one was given by Barequet and Sharira [3], where the matching algorithm detects and fits parts of the boundary edges and then fills holes with triangular patches. Another solution proposed by Kumar and Barequet [2] makes it possible to select and combine pairs of fiting edges. For each such pair the area between the corresponding edges is calculated. The edge combining occurs sequentually in ascending order according to the value of calculated area.

Simplification of the mesh topology. The solution to this problem was obtained according to the proposal of Guskov and Wood [7]. The algorithm finds and removes all tunnels in the mesh with a maximum size specified by the user. The first stage of the algorithm is to find the tunnel. This occurs if the two loops are closed triangles belonging to the mesh in connected with the same edge. In the next step the connected triangles are cut and the remaining two holes are filled with Liepa algorithm [8].

2.2 Volumetric Repair

Volumetric algorithms are newer approach for removing artifacts in 3D mesh models. The main steps that characterize these algorithms are: converting the model to an intermediate volume representation (grid points, mesh adaptation such as octal tree or BSP tree), discretization to the voxel representation with morphological operations such as dilation, erosion, smoothing, flood-fill to determine the interior part of the object), extracting the surface of the object. Detailed description of the volumetric techniques can be found in [9]. In our implementation we used volume processing for mesh merging.

2.3 Fitting the Head Shape

We assumed that the shape of the head can be approximated as analytical surface with a generalized quadric representation of an ellipsoid. The algorithm for finding such representation using the Least Squares Approach was described by Dai, Newmann and Cao [6].

The ellipsoid equation can be written in matrix form:

$$X^T A X = 1 \quad (1)$$

where:

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, A = \begin{bmatrix} 1/a^2 & 0 & 0 \\ 0 & 1/b^2 & 0 \\ 0 & 0 & 1/c^2 \end{bmatrix}$$

Taking into account the transformations, such as translation and rotation, the equation takes the form:

$$\hat{X}^T A_1 \hat{X} = 1 \quad (2)$$

where:

$$\hat{X} = \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix}, A_1 = R^T A R, R = R(\theta_x, \theta_y, \theta_z)$$

R – rotation matrix, and $\theta_x, \theta_y, \theta_z$ – rotation angles.

Only six independent components is needed to represent A_1 as the matrix is symmetric. We can expand the equation (2) and compare it with the general equation for quadric:

$$k_1x^2 + k_2y^2 + k_3z^2 + k_4xy + k_5yz + k_6zx + k_7x + k_8y + k_9z = 1 \quad (3)$$

where parameters $k_1 \dots k_9$ can be calculated using the Least Square Analysis for a given set of points selected from the scanning data.

Then the task is reduced to determining the values x_0 , y_0 , and coefficients of the rotation matrix A_1 by solving the linear equation corresponding to (2). The angles θ_x , θ_y , θ_z are computed as Euler angles regarding the rotation matrix A_1 . The ellipsoid parameters can be calculated the singular values (s_1, s_2, s_3) of matrix A_1 : $a = \sqrt{1/s_1}$, $b = \sqrt{1/s_2}$, $c = \sqrt{1/s_3}$.

3 Implementation

Application was implemented in C++ using Microsoft Visual Studio 2010 and software libraries for RapidForm.

3.1 *Strategy of Face Shape Reconstruction*

The task can be divided into several stages, which should be performed in the given sequence to achieve the best possible quality of the output model. First, some pre-processing such as model positioning and data filtration is necessary. Model positioning sets the global coordinates for different mesh grids imported from scanner. We choose the frontal position of the model, such that the axes of eyes and the line of the nose are in coincidence with the axes of the coordinate system. The artifacts and noises in scanning data are also eliminated at this stage.

Afterwards, the process of updating the model with data resulting from finding symmetry plane in face structure is performed. This process is optional and semiautomatic, the operator has to decide which part of face structure can be mirrored and copied to the output model. Usually such operation can be done for the structure of ear as for a single scan this part is weakly represented.

Next, the shape of the head should be reconstructed. We assumed for computational simplicity that the back of human head has ellipsoidal shape. However, this assumption is not always correct, it gives the approximate solution for the full head reconstruction from incomplete frontal data. The comparison of real full head scans and these reconstructions was submitted as conclusions to the article. We begin with the mesh of sphere, and then we move and do the affine transformation of this structure until we obtain the ellipsoid fitting the arbitrary chosen point data from the scan. Then we cut the resultant ellipsoid with a regression plane approximating the

border points of the scan. Finally, the face scan and the back side of the ellipsoid are merged together using a volume merging algorithms.

The resultant mesh is then finally cleaned from noises. The remaining holes are filled and artifacts that appeared at the step of merging are removed. Also the texture is remapped into the reconstructed mesh.

3.2 Mesh Modifications

Face symmetry. This operation is optional depending on the specific model. Some scans are made in the way that the elements of symmetrical faces (eg. ears) are represented better than others. Copying the mirrored parts is the way to obtain the more complete representation of the model. The following method is useful here:

- Create the mirror plane – the function `CalcMirrorPlane` from RapidForm library, basing on the PCA analysis, was used here. The mirror plane is fixed around three arbitrary chosen points lying vertically along the center of the scan grid (the nose line).
- Select items to copy – the selection is done manually by the operator.
- Copy and merge the selected data – the function `DivideAndMirrorShell` from RapidForm library is called with input parameters: the mirror plane and the selected items to be mirrored and copied.

Fitting ellipsoid to head shape. Certain number of points on the face must be selected first to create an ellipsoid, which will later be approximating the shape of the head. The position, size and orientation of such quadric is adjusted automatically using the method based on SVD described in section 2.3. The selection of points should be selected manually to ensure good accuracy of average matching and tight fitting at the boundaries of the scan. The conversion of computated ellipsoid to the mesh is implemented by resizing the template sphere mesh of a given density.

Trimming the ellipsoid mesh. The issue is divided into stages as follow:

- Create of the cutting plane by defining the regression plane for the boundary points of the scan. Sometimes such method does not give the satisfactory results and then it is good to select manually the same number of points on both sides of the face along the boundaries, at least 3 points for each side.
- Determine the intersection of the ellipsoid and the cutting plane. RapidForm library function `DivideAndMirrorShell` is useful here.
- Remove the front part of the ellipsoid mesh.

Merging the front part of the scan and the ellipsoid mesh. At the last stage the integration of the scan and the prepared ellipsoid mesh is conducted. The purpose is to obtain a representative surface model of head in form of a uniform mesh. The meshes are intergated using the volumetric merging method (sect. 2.2). The merging process is followed by surface-oriented procedures for mesh repair described in section 2.1. These methods are required not only to fill the holes resulting from

scanning errors, but also to close the gaps between the front and back of the head. It is also important to copy the texture into the resultant surface model with the high resolution textures by combining two parts.

4 Examples of Face Reconstruction

A couple of tests were performed to show the effect of face model reconstruction. The scanning process may produce unexpected results and is not free from errors. The reasonable problem, for instance, was scanning the surface of hair. Another scanning problem was reproduction of the glasses due to their optical distortion. Also, not centered position of the face relative to the scanner can cause unbalanced representation of its shape.

No reproduction of hair and ears in the scan can cause a significant reduction of the volume and flattening of the reconstructing shape of the head. In the model shown in Figure 2 a virtual shape reconstruction of the left ear was done and its mirrored representation was copied to the right side. The important advantage of the reconstruction is a good representation of forehead and it is sometimes supported by the imposition of a cap.

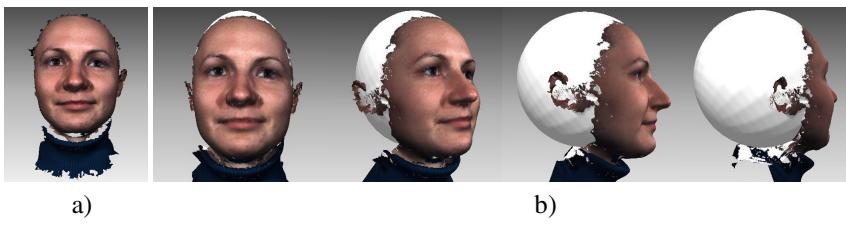


Fig. 2 An example of face reconstruction from a single scan: a) single scan, b) face reconstruction

4.1 Comparison to Complete Face Reconstruction

Three variants of the head reconstruction are compared in this section. The first variant is based on the reconstruction of the basic facial scan (fig. 1b). Another one is a scan of the head in the cap, which gives better surface coverage. The second approach gives more accurate reconstruction of almost the entire surface of the head and eliminates the measurement error caused by hair. The last variant is the reference head reconstruction base on the integration of four face scans (fig. 3). The comparison of reconstruction quality considered parameters such as: deviation of the reconstructed surface and the reference surface, surface area and volume.

The model in its first variant presents a fairly good representation of facial details (fig. 4a). Comparing the reconstruction to the reference model the significant

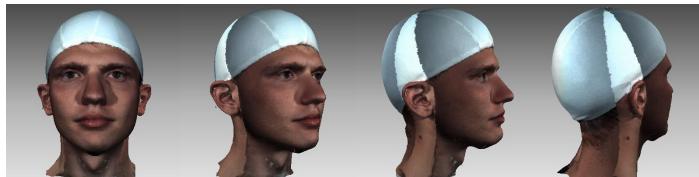


Fig. 3 Reference head reconstruction based on four scans

differences are visible. The reconstructed head model is of the smaller height, flatter at the top and smaller size (fig. 4b). The difference od the distance between farest back point in lateral view is abour 3cm. distance from the back of the head between the two models is the order of the width of the ear (about 3 cm). The surface area of the head is smaller than the surface area of the reference model of approximately 170cm^2 , and the volume is less than about 790cm^3 (see Table 1).

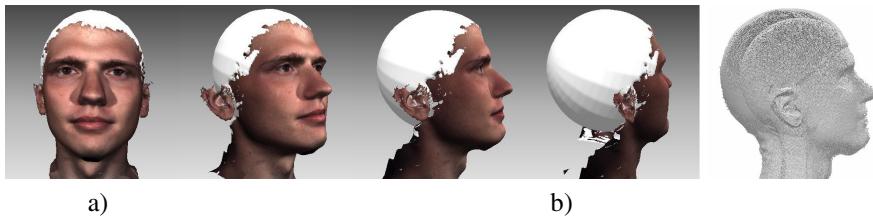


Fig. 4 a) Reconstruction of head model using basic face scan, b) comparison to reference head model



Fig. 5 a) Reconstruction of head model using face scan with cap, b) comparison to reference head model

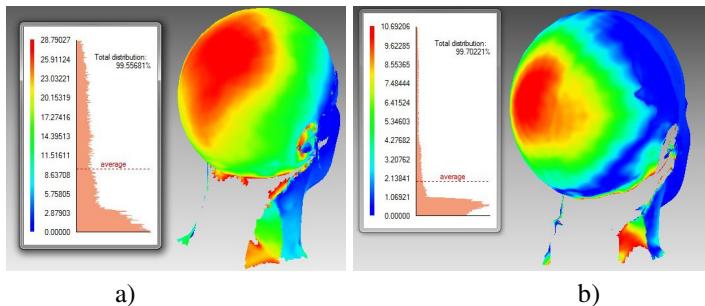
Even better results were obtained for the second reconstruction, which was performed for the face scan with cap (fig. 5a). The average difference in distance between the reconstruction and the reference model is about 2 mm and the maximal reaches 10 mm (fig. 5b). The reconstructed model is larger than the reference one. The difference in surface area is about 80 mm^2 and in volume reaches 690 mm^3 .

Table 1 The differences between the reconstructed models and the reference model

	Surface area	Volume	Avg deviation	Max deviation
1st model	-170mm ²	-790mm ³	9.5mm	30mm
2nd model	+80mm ²	+690mm ³	2mm	10mm

The second case, where the head was covered with hair cap, gives a good representation of the shape in larger area. This facilitates the surface reconstruction and makes it possible to prepare the ellipsoid more precisely, it is in fact similar to the shape of real head. representation of the shape of the head during the reconstruction and preparation of an ellipsoid with dimensions similar to the shape of the head.

The detailed deviation analysis shown in fig. 6 also confirms the above observations. The histogram of model deviations for both cases is presented.

**Fig. 6** Comparison of the deviations in the first (a) and second (b) models

5 Conclusions

It has been shown that the accuracy and the scope of the input representation increases the quality of the reconstruction, and the more detailed representation acquired in the scanning process is reflected in the quality of the final model.

The resultant models were described and compared with the complete reference model of the head formed by the merging the partial scans taken from four sides of the head using the techniques described in [10]. Such comparison shows the imperfections of reconstruction from a single scan. It may be noted that the quality of the fitting ellipsoid for a single scan is largely dependent on the representation completeness and expand the 3D scanning products. For this reason, such approach can be useful to obtain approximate and illustrative models, for example, for prototyping purposes, where solid models are necessary. Unfortunately, it is problematic to use such an approximate reconstruction for quantitative analysis, such as in medical diagnostics.

Acknowledgements. This work was supported in part by the Ministry of Science and Higher Education of Polish Government, under the research project N N516 479740.

References

1. Botsch, M., Kobbett, L., Pauly, M., Alliez, P., Lévy, B.: *Polygon Mesh Processing*. A K Peters (2010)
2. Barequet, G., Kumar, S.: Repairing CAD Models. In: Proc. 8th IEEE Conf. on Visualization (VIS), Phoenix, AZ, pp. 363–370 (1997)
3. Barequet, G., Sharir, M.: Filling Gaps in the Boundary of a Polyhedron. *Computer-Aided Geometric Design (CAGD)* 12(2), 207–229 (1995)
4. Boehnen, C., Flynn, P.: Accuracy of 3D Scanning Technologies in a Face Scanning Scenario. In: IEEE 5th Int. Conf. on 3D Digital Imaging and Modeling, pp. 310–317 (2005)
5. Borodin, P., Novotni, M., Klein, R.: Progressive Gap Closing for Mesh Repairing. In: Computer Graphic International Conference, pp. 201–213 (2002)
6. Dai, M., Newman, T.S., Cao, C.: Least-squares-based fitting of paraboloids. *Pattern Recognition* 40(2), 504–515 (2006)
7. Guskov, I., Wood, Z.J.: Topological Noise Removal. *Graphics*. In: Interface Proceedings: Ottawa, Canada, pp. 19–26 (2001)
8. Liepa, P.: Filling Holes in Meshes. In: Eurographics Symposium on Geometry Processing, pp. 200–205 (2003)
9. Nooruddin, F.S., Turk, G.: Simplification and Repair of Polygonal Models Using Volumetric Techniques. *IEEE Trans. on Visualisation and Comp. Graphics*, 191–205 (2003)
10. Skabek, K., Tomaka, A.: Automatic Mering of 3D attribute meshes. In: Kurzydowski, M., et al. (eds.) *Computer Recognition Systems 2. ASC*, vol. 45, pp. 645–652. Springer, Heidelberg (2007)

The Effectiveness of Matching Methods for Rectified Images

Pawel Popielski, Zygmunt Wrobel, and Robert Koprowski

Abstract. Medical diagnostics today is based mainly on invasive methods and it should be strongly emphasised that they include not only the X-ray imaging, but also CT and MRI scanning. For several years in various research centres, there have been attempts to create a non-invasive medical diagnostic systems based on the fusion of photogrammetric and computer vision methods. Both the complexity of the problem and commitment to used well-known methods of diagnosis in medical circles did not allow for the creation of a fully functional prototype of system that could be implemented. In the paper, the authors present the problem of 3D reconstruction with a diagnosis of suitability of various matching methods used for rectified images. The result clearly indicate the superiority of the algorithm based on variational solution. The authors in their work on the development of photogrammetric non-invasive medical diagnostic system have not come across such an analysis. Therefore, they concluded that presenting such an analysis will be useful in further research.

1 Introduction

Owing to full automation, 3D reconstruction from non-metric images in the visible band can be used in medicine, namely in the diagnosis of faulty posture, metabolic diseases, bone or plastic surgery [3, 26, 27]. Prerequisites for admission of this modern diagnostic method for use in medical practice is, among other things, immediacy of results, non-invasiveness, high accuracy of reconstructed geometry, low cost and compactness of the diagnostic installation [2, 29]. We assume that the system will be able to automatically extract geometrical characteristics what enable us to determine among others the phase of a scoliosis or degree of a overweight.

Pawel Popielski · Zygmunt Wrobel · Robert Koprowski
Institute of Computer Science, University of Silesia, 41-200 Sosnowiec, ul. Bedzinska 39
e-mail: pawel.popielski, zygmunt.wrobel,
robert.koprowski@us.edu.pl

3D reconstruction methods previously used in medicine, based on laser scanning or projection of various geometrical elements [27], did not enable to obtain sufficient measurement accuracy, reconstruction completeness and automation.

Throughout the previous century, photogrammetry [22] dealt with 3D reconstruction but finally computer vision allowed us to complete the process of full automation of 3D reconstruction from images [10, 34]. Among other things, owing to the development of 3D reconstruction algorithms, it was also possible to work with non-metric cameras (i.e. not having specific, pre-established parameters of interior orientation [16] and fully eliminated aberrations [13]).

At a certain level of abstraction, it can be assumed that the 3D reconstruction process consists of four main stages [22, 33] what will be described closely in chapter two. First, the intrinsics and aberrations must be determined. Then the extrinsics must be calculated. In next step matching process has to be done and then point's coordinates X, Y, Z have to be calculated. Generally by X, Y, Z we mean coordinates in the object's space and by x, y, z we mean coordinates in the image space. The image space coordinates are not necessarily equal to screen coordinates.

The third stage is currently the centre of attention for the authors of this paper. Detection of points [31, 32, 35] itself brings many challenges as the object must be covered with a sufficiently dense network of points to provide high accuracy of shape reconstruction [25]. The application of epipolar geometry rules enables to completely eliminate a computationally expensive problem of point detection to focus on an efficient matching process.

The section 2 presents the process of obtaining the object coordinates X, Y, Z and the matching methods used by the authors. The section 3 focuses on the results of 3D reconstruction. The section 4 includes a summary with an attempt to identify the most effective method.

2 3D Model Reconstruction

Photogrammetry [22] usually uses expensive imaging devices with intrinsics, i.e. the principal distance c_k and image coordinates of the principal point of mapping, fixed during the manufacturing process. However, cheap non-metric cameras or webcams have unknown intrinsics and lenses loaded with aberrations. Intrinsics and aberrations will be determined in the process of calibration [13].

2.1 Calibration of Imaging Device

2.1.1 Distortion

Radial and tangential distortions are the most harmful aberrations [8, 13]. The radial distortion is related to the radial shape of lenses that the lens is made of. Eliminating this error during production would require e.g. making parabolic lenses or placing

the aperture exactly in the perspective centre. According to [13], radial distortion can be characterized by the first few terms of a Taylor series expansion

$$\begin{aligned}x_{correct} &= x(1 + k_1r^2 + k_2r^4 + k_3r^6) \\y_{correct} &= y(1 + k_1r^2 + k_2r^4 + k_3r^6)\end{aligned}\quad (1)$$

where $x_{correct}$ refers to the corrected coordinates, x - the original image coordinates, k_n - radial distortion coefficients, and r - the radius from the principal point in the image. Tangential distortion is associated with non-parallel placement of lenses with respect to each other and to the image plane. According to [7], tangential distortion can be characterised by minimum two parameters, p_1 and p_2

$$\begin{aligned}x_{correct} &= x + [2p_1y + p_2(r^2 + 2x^2)] \\y_{correct} &= y + [p_1(r^2 + 2y^2) + 2p_2x]\end{aligned}\quad (2)$$

where $x_{correct}$ and x are the coordinates, and r is the radius from the principal point. The five parameters obtained in this way create a distortion vector. According to [8], using homographic transformation and having an image of a flat object, whose object coordinates are known, such as the chessboard, it is possible to find distortion coefficients by solving a system of equations with five unknowns. With the coordinates of four extreme corners, 2×4 equations are obtained, which enables to calculate the five-element vector of distortion.

2.1.2 Intrinsic

Projective transformation mapping of the object space into image space can be written as the product of the matrix [16], in which M is introduced as the matrix of the parameters of interior orientation, or in short, the interior orientation matrix

$$q = |MQ| \quad \text{where} \quad q = \begin{vmatrix} x \\ y \\ w \end{vmatrix}, \quad M = \begin{vmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{vmatrix}, \quad Q = \begin{vmatrix} X \\ Y \\ Z \end{vmatrix} \quad (3)$$

The parameters of the matrix M are the interior orientation elements i.e. a focal length f and coordinates of the principal point c . Homogeneous coordinates in the image space q are a result of coordinate transformation in the object space Q by the interior orientation matrix M . This gives a system of equations with four unknowns. When examining the point in object and image spaces, it appears from pure rules of geometry [36, 37] that

$$P_{img} = R(P_{obj} - T) \quad (4)$$

where P_{img} refers to the coordinates in the image, P_{obj} means the object coordinates, R is the matrix of rotation, and T is the translation vector between the image and object systems. Matrices R and T are called the exterior orientation parameters or extrinsics. Since the rotation consists of three angles and shift by Translation ternary vector, the result is a system of equations with six unknowns. After connecting the

current system of equations (4) with the equations resulting from (3), the result is a system of equations with ten unknowns. Using the chessboard image again, where due to homography only the four outermost points are effective, it can be observed that one image is not enough to solve the system of equations. Assuming the availability of $K - \text{images}$, where each image has $N - \text{corners}$, the result is $2NK$ systems of equations. Multiplied by two, because in the image space there are two dimensions. It follows that for every image, six parameters of exterior orientation need to be calculated. Next four parameters of the interior orientation are the same for all images. The result is the following equation

$$2NK = 6K + 4 \quad (5)$$

Bearing in mind that for our chessboard $N = 4$, the result is $K = 2$. Taking into account random errors and numerical stability of the equations system, it is clear that more images are needed. Practical experience shows that an accurate calibration is obtained with at least fifteen images. The result is a distortion vector and matrices of interior orientation parameters for each camera.

2.1.3 Extrinsic

To calculate the exterior orientation parameters, the point P located in the object space needs to be expressed by means of coordinates of the left image

$$P_l = R_l P + T_l \quad (6)$$

and of the right one

$$P_r = R_r P + T_r \quad (7)$$

where P_l refers to the coordinates of the point P in the left image system, R_l is the rotation matrix of the object system to the left image system, and T_l is the vector of translation of the object system to the left image system; the symbols for the right image are analogous. These values are a byproduct of the process of calculating the interior orientation parameters. At the same time, based on the geometric relationship, the coordinates of the left image system are a function of the right image system coordinates according to the following formula

$$P_l = R^T (P_r + T_r) \quad (8)$$

Where R is the sought rotation matrix of the right image with respect to the left image and T is the vector of translation of the right image into the left one. Solving the equations (6),(7) and (8) due to R and T , the result is

$$\begin{aligned} R &= R_r R_l^T \\ T &= T_r - R T_l \end{aligned} \quad (9)$$

that is, the sought exterior orientation parameters.

2.2 Epipolar Geometry and Relative Orientation

Epipolar geometry [12, 15, 34] combines two images of the same scene by introducing the concept of epipolar plane. Epipolar geometry introduces the concepts of the Essential matrix E and the Fundamental matrix F which connect the points of the left image with epipolar lines in the right image. The Essential matrix E [15]

$$P_r^T EP_l = 0$$

is expressed in units of the object system or units of the image system, that is, in meters or millimetres. Whereas the Fundamental matrix F [15]

$$q_r^T F q_l = 0$$

is expressed in pixels, owing to the introduction of the intrinsics matrix M to the Essential matrix E . According to the rules of epipolar geometry, an epipolar line p_r in the right image corresponds to the image of the point P_l in the left image. This is due to a defect in both matrices, E and F , since their rank is always equal to $n - 1$. Based on the epipolar condition, pairing of homologous points, that is matching, can be simplified to the task in 1D space. The obvious fact follows from this condition, namely the equivalent of the point in the left image is exactly on one line in the right image. In addition, assuming that all of the epipolar lines in both images are horizontal and the corresponding epipolar lines of the two images are on the same horizontal line, much lower computational cost of pairing process is required. Bringing the two images to such a position is called an rectification or relative orientation. Of course, the rectification can be achieved in two ways. When taking images, video cameras or cameras can be set in such a way so that the optical axes of lenses are perfectly parallel to each other (horizontal) and image planes are in the same vertical plane. If it is not possible to provide such an arrangement of imaging devices during exposure, the taken image can be brought into such a position by using appropriate transformations. In the study, the authors carried out a process of rectification using Bouguet's algorithm [6], which minimises image distortion, while maximising the common area. To achieve it, rotation of matrix R is divided evenly into two matrices r_l and r_r . To obtain the horizontal position of the epipolar lines in the left image, the image must be transformed in such a way that the epipole turns into point at infinity. It can be achieved by multiplying by the matrix Rectify equal to

$$R_{rectify} = \begin{vmatrix} e_1 \\ e_2 \\ e_3 \end{vmatrix} \quad (10)$$

where

$$\begin{aligned} e_1 &= \frac{T}{\|T\|} \\ e_2 &= \frac{|-T_y \ T_x \ 0|^T}{\sqrt{T_x^2 + T_y^2}} \\ e_3 &= e_1 \times e_2 \end{aligned}$$

Then the epipolar lines in the right image are brought to a horizontal position, corresponding to the epipolar lines in the left image, by rotating the two images by

$$\begin{aligned} R_l &= R_{rectify} r_l \\ R_r &= R_{rectify} r_r \end{aligned}$$

The calculated matrices enable reprojection of the images, which results in rectified images. Having images that are already calibrated and rectified the process of matching along epipolar lines can be started.

2.3 Matching

In earlier authors studies, when reconstructed images were not subjected to rectification, it was necessary to perform detection of points before matching. The methods of Harris [14] and Noble [28] demonstrated the highest efficiency. Then matching was performed using a fast normalised cross-correlation method [23]. The process, however, did not provide sufficient performance for real-time applications [19, 20].

Computationally expensive process of detection of points was eliminated by image rectification. The points are, in sequence, chosen in the left image and then their corresponding points in the right image are searched with a chosen method. The search starts on the right image from the position x_l - the point coordinate in the left image, moving towards the left. Since the images are rectified, the coordinate $y_l = y_r$. When the function of matching gives a sufficiently high value, then the current position in x_p is stored. Then follows the calculation of the transverse parallax also called disparity vector $p = x_l - x_p$, which is used to calculate the image coordinates in the process of triangulation.

Each matching method uses image pre-processing to improve its quality, for example, by using the normalization function. The result of the matching function is checked by eliminating incorrect pairing by building the fundamental matrix [24] using the estimation algorithm RANSAC [11]. The authors carried out estimation with a very strict parameter $t = 0.001$, which determines the distance of points from the surface of the model. To increase the resolution of the cloud of points, the enrichment of homogeneous surface (e.g. the human body) was applied, by projection of the high-contrast pattern [9, 30]. Pattern composition method is based on the studies of Bailey and Borwein [4, 5] which prove the statistical nature of the Ludolph's number. With the quasi-normal nature of the number π , uniqueness of the pattern is ensured.

2.3.1 Simple Block Matching (SAD)

In [18], matching is performed by moving in the right image a small SAD window(sum of absolute difference) sized from 5×5 to 21×21 pixels. The SAD function responds to points with a strong contrast. Therefore, scenes full of details are well correlated. However, uniform surfaces, devoid of details, are mapped as a hole

in a cloud of points. To prevent this, for this method and for the other authors use projection of an artificial pattern consisting of black and white elements.

2.3.2 Semiglobal Matching (SGBM)

Based on [17], the authors utilised pixelwise, Mutual Information based matching cost for compensating radiometric differences of input images. Pixelwise matching is supported by a smoothness constraint that is usually expressed as a global cost function. SGBM method performs a fast approximation by pathwise optimizations from all directions. In single-pass version, only 5 directions are considered. In another implementation (F-SGBM), eight directions are considered, just to conduct the comparison of time-consumption and cloud resolution to results of a single-pass algorithm.

2.3.3 Variational Matching Algorithm (VF)

Based on [21], multi grids were implemented in combination with the multi-level adaptation techniques, where heterogeneous adaptive structures can be combined with a variational solution for the parallax vector estimation in real time.

2.4 *Triangulation*

Assuming that the images are rectified and the disparity vector p is known, it is possible to start the calculation of the object coordinates. According to [22], the coordinates are calculated from the following formulas

$$X = \frac{Tx}{p} \quad Y = \frac{Ty}{p} \quad Z = \frac{Tf}{p}$$

Since these equations have been derived from the similarity of triangles, the method is called triangulation.

3 Test Procedure

The test rig consists of a projector and two webcams Logitech C920 HD Pro Webcam. The used sensors enable to obtain an image with a resolution of 1920 by 1080 pixels. Tessar lens designed by Carl Zeiss is characterized by the elimination of most of the optical errors. However, only a partial elimination of distortion was possible. The webcams were placed on opposite ends of the 30 cm base.

The accuracy of 3D reconstruction increases with the length of the shooting base, keeping in mind, of course, the condition of the base-depth ratio [22]. Projection of the artificial pattern helps to enrich homogeneous surfaces.



Fig. 1 Photos and theirs 3D reconstructions with VF

The reconstruction was carried out for three different scenes in succession for four matching algorithms without and with projected pattern. The third and fifth column (Table 1) shows the number of points forming a cloud. Since the authors' aim is to implement a system running in real-time, the operating time of individual algorithms was also measured.

Table 1 Stereo correspondence result

Picture	Pairing method	Reconstruction without pattern		Reconstruction with pattern	
		Number of points in the cloud	Time (ms)	Number of points in the cloud	Time (ms)
Man	SAD	66 463	48	62 965	68
	SGBM	182 904	352	215 671	291
	F-SGBM	175 258	390	204 044	388
	VF	429 045	1205	460 070	1191
Fern	SAD	40 156	48	55 207	71
	SGBM	225 379	422	236 139	423
	F-SGBM	208 348	428	221 201	504
	VF	442 565	1300	455 558	1198
Arrangement	SAD	38 023	70	34 531	71
	SGBM	214 479	298	244 080	298
	F-SGBM	189 686	547	224 454	551
	VF	443 253	1195	453 025	1292

4 Conclusions

The obtained results (Table 1) confirm high effectiveness of matching based on the VF. Achieved points cloud (Fig. 1) easily enables to apply further geometrical analysis. Clearly some more work is needed to tune up the VF algorithm to achieve better time effectiveness. Fast block matching based on SAD window is by far the least effective when confronted with the task of the human body reconstruction.

All matching algorithms are very sensitive to the scene illumination and even small changes in the amount of light may produce quite different number of disparities.

It can be observed that some points have not been matched, because they were not visible in both images at the same time.

In further work, the authors will focus on developing VF algorithm. Tuned algorithm will be able to significantly push up the finalisation and introducing of the photogrammetric non-invasive medical diagnostic system.

References

1. Daniilidis, K., Klette, R.: Imaging Beyond the Pinhole Camera", Computational Imaging and Vision, vol. 33. Springer (2006)
2. D'apuzzo, N.: Automated Photogrammetric Measurement of Human Faces. In: Int. Archives of Photogrammetry and Remote Sensing, Hakodate, Japan, vol. XXXII, Part B5, pp. 402–407 (1998)
3. D'apuzzo, N.: Measurement and modelling of human faces from multi images. International Archives of Photogrammetry and Remote Sensing 34(5), 241–246 (2002)
4. Bailey, D., Borwein, J., Mattingly, A., Wightwick, G.: The Computation of Previously Inaccessible Digits of and Catalan's Constant, Notices of the American Mathematical Society (2011),
<http://crd.lbl.gov/~dhbailey/dhbpapers/bbp-bluegene.pdf> (accessed April 15, 2011)
5. Berggren, L., Borwein, J.M., Borwein, P.B.: Pi: a Source Book. Springer, New York (2004)
6. Bouguet, J-Y.: Camera calibration toolbox for Matlab,
http://www.vision.caltech.edu/bouguetj/calib_doc/index.html
7. Brown, D.C.: Decentering Distortion of Lenses. Photometric Engineering 32, 444–462 (1966)
8. Brown, D.C.: Close-range camera calibration. Photogrammetric Engineering 37, 855–866 (1971)
9. Chang, Y.: A Photogrammetric System for 3D Reconstruction of a Scoliotic Torso, A Master Thesis, Department of Geomatics Engineering, University of Calgary, Canada (2008)
10. Cyganek, B., Siebert, J.: An Introduction to 3D Computer Vision Techniques and Algorithms. Willey (2009)
11. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM 24, 381–395 (1981)

12. Forsyth, D., Ponce, J.: Computer Vision: A Modern Approach. Prentice-Hall (2003)
13. Fryer, J.G., Brown, D.C.: Lens distortion for close-range photogrammetry. Photogrammetric Engineering and Remote Sensing 52, 51–58 (1986)
14. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of Alvey Vision Conference, vol. 15, pp. 147–151 (1988)
15. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2006)
16. Heikkila, J., Silven, O.: A four-step camera calibration procedure with implicit image correction. In: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition, p. 1106 (1997)
17. Hirschmuller, H.: Stereo Processing by Semiglobal Matching and Mutual Information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2008)
18. Konolige, K.: Small vision system: Hardware and implementation. In: Proceedings of the International Symposium on Robotics Research, Hayama, Japan, pp. 111–116 (1997)
19. Korzynska, A., Iwanowski, M.: Multistage morphological segmentation of brightfield and fluorescent microscopy images. Opto-Electronics Review 20(2), 174–186 (2012)
20. Korzynska, A., Hoppe, A., Strojny, W., et al.: Investigation of a combined texture and contour method for segmentation of light microscopy cell images. In: Proceedings of the Second IASTED International Conference on Biomedical Engineering 2004, pp. 234–239 (2004)
21. Kosov, S., Thormählen, T., Seidel, H.-P.: Accurate Real-Time Disparity Estimation with Variational Methods. In: 5th International Symposium on Visual Computing, USA (2009)
22. Kraus, K.: Photogrammetry. Walter de Gruyter, Berlin (2007)
23. Lewis, J.P.: Fast normalized cross-correlation. Vision Interface, 120–123 (1995)
24. Luong, Q.T., Faugeras, O.D.: The Fundamental Matrix: Theory, Algorithms, and Stability Analysis. International Journal of Computer Vision 17(1), 43–75 (1996)
25. Malian, A., Azizi, A., Van Den Heuvel, F.A.: Medphos: A new photogrammetric system for medical measurement. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35(B5), 311–316 (2004)
26. Mitchell, H.L.: Applications of digital photogrammetry to medical investigations. ISPRS Journal of Photogrammetry and Remote Sensing 50(3), 27–36 (1995)
27. Mitchell, H.L., Newton, I.: Medical photogrammetric measurement: overview and prospects. ISPRS Journal of Photogrammetry and Remote Sensing 56(5-6), 286–294 (2002)
28. Noble, A.: Descriptions of Image Surfaces, PhD thesis, Department of Engineering Science, Oxford University (1989)
29. Patias, P.: Medical imaging challenges photogrammetry. ISPRS Journal of Photogrammetry and Remote Sensing 56(5-6), 295–310 (2002)
30. Popielski, P., Wróbel, Z.: The feature detection on the homogeneous surfaces with projected pattern. In: Piętka, E., Kawa, J. (eds.) ITIB 2012. LNCS, vol. 7339, pp. 118–128. Springer, Heidelberg (2012)
31. Porwik, P., Para, T.: Some handwritten signature parameters in biometric recognition process. In: Proceedings of the ITI 2007 29th International Conference on Information Technology Interfaces Book Series: ITI 2007, pp. 185–190 (2007)
32. Porwik, P., Wrobel, K., Doroz, R.: The Polish Coins Denomination Counting by Using Oriented Circular Hough Transform. In: Kurzynski, M., Wozniak, M. (eds.) Computer Recognition Systems 3. AISC, vol. 57, pp. 569–576. Springer, Heidelberg (2009)
33. Schenk, T.: Digital photogrammetry, TerraScience, Laurelville, Ohio, 428 (1999)
34. Shapiro, L., Stockman, G.C.: Computer Vision. Prentice-Hall (2002)

35. Wróbel, K., Doroz, R.: The new method of signature recognition based on least squares contour alignment. In: International Conference on Biometrics and Kansei Engineering, pp. 80–83 (2009)
36. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the 7th International Conference on Computer Vision, Corfu, pp. 666–673 (1999)
37. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 1330–1334 (2000)
38. Kajan, E.: Information technology encyclopedia and acronyms. Springer, Heidelberg (2002)
39. Broy, M.: Software engineering – From auxiliary to key technologies. In: Broy, M., Denter, E. (eds.) *Software Pioneers*. Springer, Heidelberg (2002)
40. Che, M., Grellmann, W., Seidler, S.: *Appl. Polym. Sci.* 64, 1079–1090 (1997)
41. Ross, D.W.: Lysosomes and storage diseases. MA Thesis. Columbia University, New York (1977)

The Print-Scan Problem in Printed Steganography of Face Images

Włodzimierz Kasprzak, Maciej Stefańczyk, and Jan Popiółkiewicz

Abstract. The use of steganography and watermarking techniques for a secure identification and automatic authentication of the holder's photo in an ICAO-consistent (travel) document is proposed. A specific distortion of the hidden watermarks is caused by the print-scan process, as a printed photo is scanned to constitute the reference pattern in face verification. The main distortions induced by this process are identified - image rotation and cropping, nonlinear intensity change and additive color quantization noise. Three steganography methods are selected and adapted for this application. They provide hidden watermark insertion in the image domain, the Fourier-Mellin transform domain or in a net of normalized triangle regions. Test results induce the conditions for the proper use of these methods.

1 Introduction

Image steganography and watermarking are related techniques to hide information in images or mark (annotate) images with the purpose of secure transmission of hidden information or of securing the carrier image [3], [9]. Our goal is to provide additional security for photos of document holders while annotating them with hidden information (Fig. 1). This task seems to be related to both techniques. We need to hide and secure the verifying information in the image (which is steganography) but we also need to annotate the photo with a message related to the particular document (i.e. the personal data of the document's holder, the serial number of the document).

Włodzimierz Kasprzak · Maciej Stefańczyk · Jan Popiółkiewicz

Institute of Control and Computation Engineering, Warsaw University of Technology,
ul. Nowowiejska 15/19, 00-665 Warszawa, Poland

e-mail: {W.Kasprzak, M.Stefanczyk}@elka.pw.edu.pl

In the considered application case the inserted watermark/stego-object may be modified and even distorted as the carrier image goes through the print-scan (PS) process, i.e. the image is printed onto the document (onto a paper or plastic card) and the document is scanned in the course of document holder verification procedure. Several steganography methods, proposed so far, already show some resistance against the print-scan process. The adaptation to geometric transformations can be achieved by inserting the hidden information in the domain of the Fourier-Mellin transform (DFM) [12]. When the inserted stego-object or watermark has a fully symmetric pattern in the amplitude spectrum then the log-polar transform step of DFM can even be avoided [13]. Another interesting approach is the insertion of hidden information in the neighborhood of characteristic feature points detected in the image. Before watermark insertion the image region is normalized (made invariant to geometric transformations) [2]. Some methods insert the hidden information in the half-tone images created for direct printing. For example, the half-tone dots are shifted depending on the coded information [11], [4].

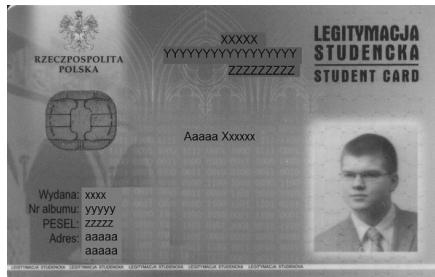


Fig. 1 The problem: to secure the authenticity of the photo of document holder

Steganography techniques differ by many aspects. We consider three crucial criteria: the domain of stego-object insertion (image space, frequency space, DFM domain, normalized image regions), the requirement of an additional secure key (e.g. [7]) or not, and the type of image synchronization between the insertion and detection steps - additional templates in the space and frequency domains or self-synchronization ability. In this paper we propose and test steganography techniques that are representative for these different criteria and which test to be sufficiently resistant against distortions introduced by the "print-scan" process.

The paper is organized as follows. Section 2 motivates and introduces our approach. Section 3 presents the analysis of distortions caused by the print-scan process. In section 4 three steganography/watermarking techniques are proposed. Some test results of the PS distortions and of our implementation are provided in section 5.

2 The Application Scenario

In Fig. 2 the scheme of our image steganography/watermarking process is presented, in which the "print-scan" process (PS) distorts the image and makes it difficult for the receiver to detect the hidden (verifying) information. Hence, there is an additional stage needed - the cancellation of distortions caused by printing and scanning the "paper-like" identity document. Eventually, this may also require to provide (explicit or implicit) *synchronization* (i.e. allowing proper image localization) between sender and receiver.

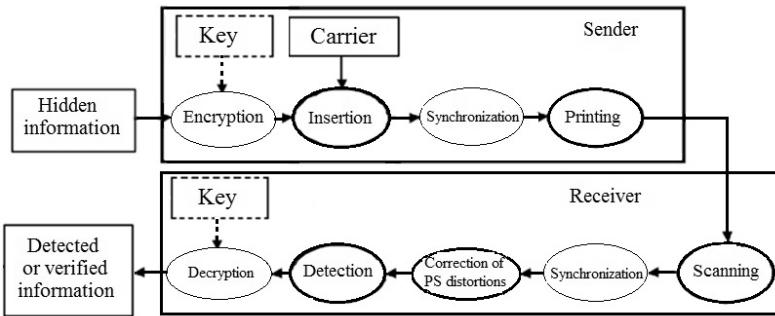


Fig. 2 The steganography/watermarking -based authentication process for a printed photo of document holder

The insertion of hidden information into the carrier image can be decomposed into following steps: 1) Coding (encryption) of hidden information; 2) Insertion of the steganography object; 3) (option) Addition of synchronization marks. At the receiver side one can distinguish following steps: 1) (option) Synchronization; 2) Correction of distortions caused by the print-scan process; 3) Detection of the steganography object/hidden watermark; 4) Decoding (decryption) of hidden information.

3 The Print-Scan Process

3.1 Distortion Models

There are two main approaches to the analysis of distortions caused by the print-scan process: the black-box approach and device modeling.

The Black-Box Approach. In one approach the distortion process is seen as a black-box and its analysis is reduced to a comparison between the original "sent" image and the "received" image. The image differences are modeled by three

transformation types [14]: image space transformations, like rotation and cropping, nonlinear pixel intensity change, and color noise.

The practical comparison results for two images related by the print-scan process can be best expressed in terms of their frequency-domain representation (DFT). The PS process deteriorates the high frequency components stronger than the mid-range frequencies, while the low-frequency components are retained in a best way. Unfortunately, the steganography object can not be hidden in the low frequencies as it will become visible in the carrier image. High-valued amplitudes of DFT coefficients are more resistant against the PS process than the low-valued ones [6].

In [14] it has been observed that the phase-difference between neighbor frequency bands is well retained in the PS process for coefficients with high amplitude values. Hence, an alternative DFT-based method, called DQIM (differential quantization index modulation) [14], is hiding the information in the phase difference of neighbor frequency coefficients.

Device Modeling. The print-scan process analysis can also be focused on individual distortions introduced by particular devices. There are typical image distortions generated by a printer: 1) gamma tweaking - a nonlinear change of the pixel intensity function, usually modeled as $I' = I \cdot \exp(r)$; 2) halftone print - a pixel is converted to a binary raster of color coefficients (which leads to a color quantization noise); 3) dot gain - the raster print leads to a phenomenon of an image looking darker than it is in reality.

The scanning device also generates several image distortions: 1) unwanted rotation of the image due to improper positioning of the document; 2) the cropping of image from the entire page, leading to unwanted image size scaling and shift; 3) gamma correction - nonlinear correction of amplitude, e.g. $I' = I \cdot \exp(g)$, where g is set according to image analysis results; 4) filtering of color quantization noise.

4 Steganography Methods

We have tested the influence of the print-scan process onto three steganography methods which differ by all the three considered criteria: the insertion domain, the synchronization mode, and whether a secure key is needed.

4.1 Synchronization

The explicit image synchronization markers are needed when the insertion method has no self-synchronization ability, i.e. is not able to adapt to effects of wrong document positioning and image cropping errors. For the first proposed method it is helpful to add markers in the image corners that define the image boundary (Fig. 3(a)). There exist also propositions to add image templates in the frequency domain [1]. In other methods the synchronization is performed automatically due to the detection of characteristic image points (Fig. 3(b)).

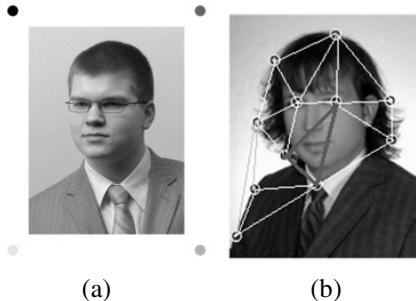


Fig. 3 Image synchronization: (a) by printing an additional frame, (b) by detecting characteristic image points

4.2 Selected Implementations

A "printed steganography" method was proposed by Fujitsu [5]. It can be characterized as follows: insertion in image blocks in the image space, explicit synchronization is needed, no key is needed for detection.

The insertion in the DFM domain was proposed in [12]. The characteristics: insertion in a transformed domain, potentially no synchronization is needed, the knowledge of character codes is needed. The insertion in the domain of the Fourier-Mellin transform makes the watermark resistant to geometric transformations of a cyclic pattern. Still an image cropping can lead to distortions, as the image is not cyclic. In principle for cyclic patterns, a shift in the image domain makes no change of the magnitude in the DFT domain. Scaling and rotation in the image domain lead to a rotation and magnitude change in the polar form of the DFT domain. The logarithm operation applied to polar DFT coefficients converts both rotation and scaling operations into a shift in the Log-Polar domain. A second DFT applied to the Log-Polar domain leads to a representation in which the magnitude values remain unchanged under all the geometric operations.

The watermark insertion in a network of triangle image regions was proposed in [2]. A short characteristics: insertion in characteristic and normalized image regions, self-synchronization is provided, a key is needed and the method provides only a verification of the watermark. It should be added that a pseudo-randomly generated watermark (with the use of a secure key) is distributed over all suitable triangle regions. The insertion process is illustrated in Fig. 8. The detection process tries to re-detect the regions (they depend on detected image corners) and to compare the extracted watermark with the known information. If sufficient number of triangles is found with positive verification result the watermark is assumed to be verified. Our modified approach uses Shi-Tomasi corner detector (instead of the Harris operator) and the Wiener filter is replaced by a smoothing filter.

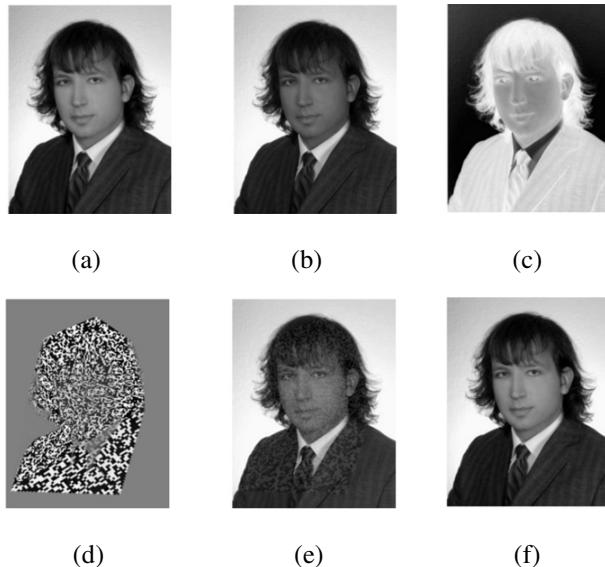


Fig. 4 The watermark insertion into normalized triangle regions freely distributed in the image: a) carrier image, b) selected color channel for information insertion, c) weight mask for insertion (high variability regions are mostly used), d) the hidden information combined with the mask, e) color plane with inserted information, f) carrier image with inserted information ($PSNR = 35dB$)

5 Tests

5.1 Testing the Distortions

Our laboratory test were performed with the use of color laser printer of type Konica Minolta. The crucial printing parameters were selected as follows:

$$\begin{aligned} 300 \text{ PPI} &\rightarrow 413 \times 551 \text{ pixel} \rightarrow 3.50 \times 4.67 \text{ cm} \\ 600 \text{ PPI} &\rightarrow 827 \times 1102 \text{ pixel} \rightarrow 3.50 \times 4.67 \text{ cm} \end{aligned}$$

A conventional scanner of type Epson Stylus was used. The document page was scanned with resolution 2400 dpi, and the photo image was cropped and scaled to the required size. The default gamma correction coefficient was 2.2. The differences of the intensity value distribution in the original digital image and the reconstructed digital image are illustrated in Fig. 5.

Let us assume here that we are able to provide a perfect synchronization between the input and output images of the PS process. Thus the only distortions will be due to linear and nonlinear amplitude changes. In order to characterize the sensitivity to such distortions we have tested separately image components in the image and frequency domains. A color image is decomposed into color components and further

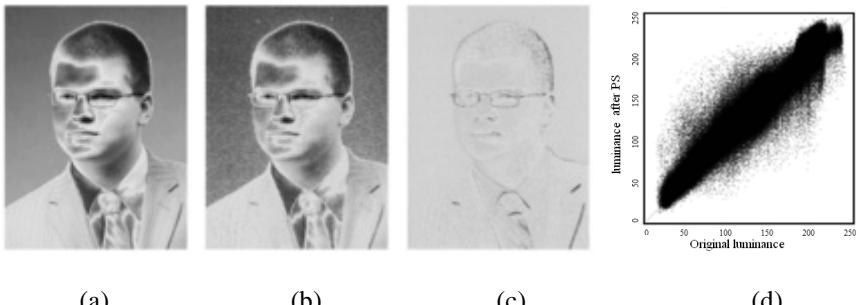


Fig. 5 Example of intensity distortion: (a) original image, (b) scanned image, (c) the relative difference image, (d) intensity changes shown on a phase plot

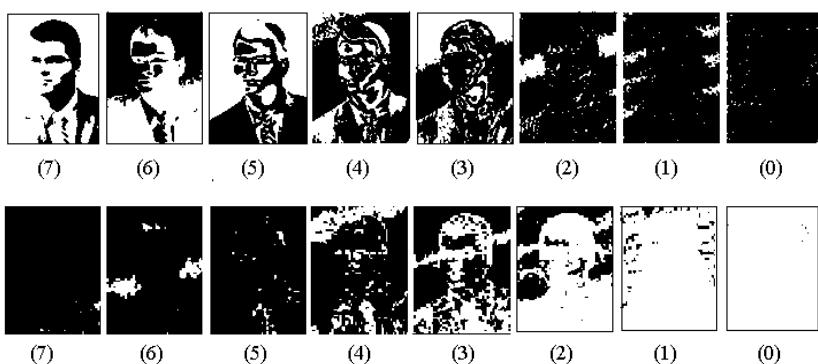


Fig. 6 (a) The image decomposition into base color components and 8-bit planes per base color (the red color component is shown here). (b) The complexities of blocks in the bit plane images - higher complexity corresponds to higher color amount

into bit planes, corresponding to the binary amplitude code (Fig. 6(a)). Each bit plane is split into blocks of size 8x8 and the "complexity" of blocks is computed as the number of bit changes along the rows and columns relative to all possible changes. A value of 0.3 is assumed to be a threshold for "sufficient" block complexity, i.e. into a sufficiently complex block a hidden information can be inserted (Fig. 6(b)). The bit planes with indices from 0 to 2 are best suitable to accept the hidden watermark, as there are around 50% of blocks of sufficient complexity. In bit planes 3-4 this amount is low, while in the bit planes 5-7 such blocks appear incidentally only.

The drawings in Fig. 7 present the statistics of bits properly reconstructed in the output image. We observe that bit planes from 0 to 4 (i.e. having the largest amount of complex blocks) have a detection rate of around 50% only (drawing (a)). When individual blocks are considered (drawing (b)), it can be observed that sufficiently complex blocks in the bit planes 5-7 have a high detection rate (> 60%).

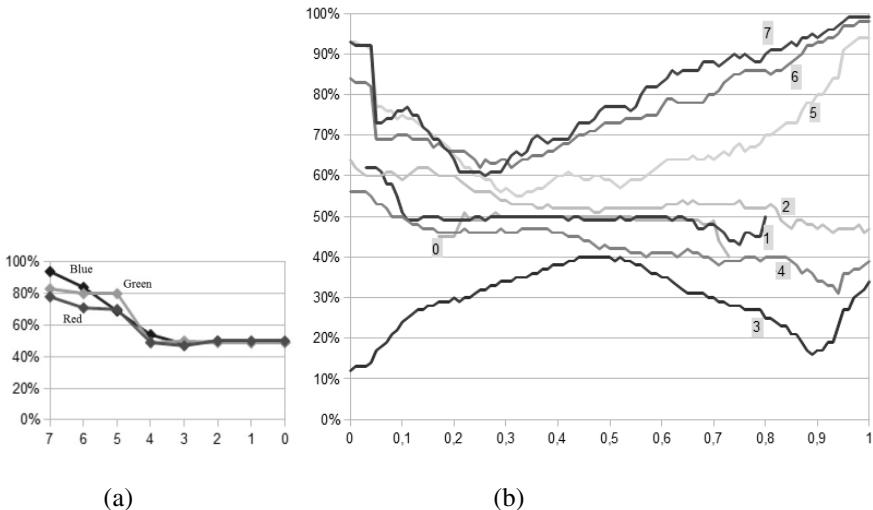


Fig. 7 The percentage of properly detected bits in the output image of the PS process with perfect geometry synchronization: (a) distribution of detection rate per bit plane, (b) detailed distribution of detection rate per plane and complexity of block

The above observations motivate us to select image blocks, showing sufficiently high variability of amplitude, instead of single bits for hidden watermark insertion, when it is performed in the image space.

5.2 Testing the Steganography Method

Some test results of our modified method of normalized triangles are summarized in Table 1. The information has been inserted in a selected color channel and major distortions caused by the print-scan process has been simulated. Insertion strength α is the magnitude coefficient of inserted code: $OUT = IN + \alpha \cdot CODE$. Two types of noise are distorting the image: an additive Gaussian noise with $\sigma = 5$ that makes image blurring, and a multiplicative noise, uniformly distributed in ± 0.03 , that simulates the quantization noise. The color change is simulated by a linear change of components in the HSV color space. The Fig. 8 illustrates an example of watermark verification - proper hidden code has been detected for majority of the triangles, while the fake code has been rejected for all of the triangles.

Table 1 Percentage of properly verified triangles for two color channels and two insertion strengths under different distortions

Parameters		Detection rates								
Channel	Insertion strength	Original image	Rotated by 5 deg	With noise	Hue change -25%	Hue change +25%	Saturation -10%	Saturation +10%	Intensity -10%	Intensity +10%
B (R G B)	0.02	100%	82 %	45 %	100 %	100 %	82 %	93 %	92 %	50 %
B (R G B)	0.05	100%	82 %	77 %	38 ¹ %	72 %	38 %	85 %	70 ¹ %	0 %
U (Y U V)	0.02	100%	82 %	54 %	64 ¹ %	71 %	50 %	85 %	100 ¹ %	82 %
U (Y U V)	0.05	100%	82 %	20 %	12 ¹ %	85 ¹ %	36 %	69 %	85 ¹ %	42 %

¹ The watermark becomes visible.

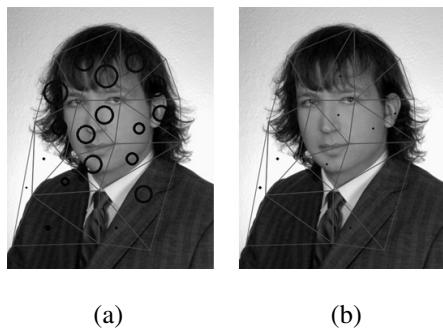


Fig. 8 Illustration of watermark detection: (a) properly verified triangles (marked by large circles) if a proper watermark has been inserted, (b) all triangles were rejected (marked by dots inside of triangles) when a faked watermark has been inserted

6 Summary

We have analyzed and tested the application of steganography methods for the verification of the identification photo, printed on an ICAO-compatible travel document. Tests performed in laboratory conditions (for a small number of images, sometimes under a perfect synchronization process) are promising. The "printed Fujitsu" method and the DFM method can save a small amount of hidden information if sent through the print-scan process under strong synchronization requirement. The "insertion in normalized triangle regions" can verify the authenticity of hidden information and it requires no external image synchronization. The methods are going to be tested now on data bases obtained in real PS processes.

Acknowledgements. The Authors gratefully acknowledge the support of the BioPKI project by the NCBiR Agency, Warsaw, under grant O ROB 0027 01/ID 27/2.

References

1. Alattar, A.M., Meyer, J.: Watermark re-synchronization using log-polar mapping of image autocorrelation. In: ISCAS 2003, Proceedings of the 2003 International Symposium on Circuits and Systems, vol. 2, pp. 928–931 (2003)
2. Bas, P., Chassery, J.M., Macq, B.: Geometrically invariant watermarking using feature points. *IEEE Trans. Image Process.* 11(9), 1014–1028 (2002)
3. Blackledge, J.: Cryptography and Steganography. New Algorithms and Applications. Center for Advanced Studies, WUT Warszawa (2011)
4. Fu, M.S., Au, O.C.: Data hiding watermarking in halftone images. *IEEE Trans. Image Process.* 11(4), 477–484 (2002)
5. N N Fujitsu Laboratories Printable Steganography (2004),
<http://jp.fujitsu.com/group/labs/techinfo/techguide/list/steganography.html>
6. Kang, X., Huang, J., Zeng, W.: Efficient general print-scanning resilient data hiding based on uniform log-polar mapping. *IEEE Trans. Information Forensics and Security* 5(1), 1–12 (2010)
7. Kasprzak, W., Cichocki, A.: Hidden Image Separation From Incomplete Image Mixtures by Independent Component Analysis. In: 13th Int. Conf. on Pattern Recognition, ICPR 1996, vol. II, pp. 394–398. IEEE Computer Society Press, Los Alamitos (1996)
8. Lin, C.Y., Chang, S.F.: Distortion modeling and invariant extraction for digital image print-and-scan process. In: Int. Symposium Multimedia Information Processing (December 1999)
9. Lipiński, P.: Watermarking software in practical applications. *Bulletin of the Polish Academy of Sciences Technical Sciences* 59(1), 21–25 (2011)
10. Nikolaidis, A., Pitas, I.: Robust watermarking of facial images based on salient geometric pattern matching. *IEEE Trans. Multimedia* 2, 172–184 (2000)
11. Rosen, J., Javidi, B.: Hidden images in halftone pictures. *Appl. Opt.* 40(20), 3346–3353 (2001)
12. O’Ruanaidh, J., Pun, T.: Rotation, scale, and translation invariant spread spectrum digital image watermarking. *Signal Processing* 66(3), 303–318 (1998)
13. Solachidis, V., Pitas, I.: Circularly Symmetric Watermark Embedding in 2D DFT Domain. *IEEE Trans. Image Processing* 10(11), 1741–1753 (2001)
14. Solanki, K., Madhow, U., Manjunath, B.S., Chandrasekaran, S., El-Khalil, I.: Print and Scan Resilient Data Hiding in Images. *IEEE Trans. Information Forensics and Security* 1(4), 464–478 (2006)

Phototool Geometry Verification

Jarosław Zdrojewski and Adam Marchewka

Abstract. This article is presenting films geometrical stability control process and its limitations in the exposure machines used in the PCB production. High imaging precision requirement is leading to sophisticated methods of thermal stability inspection. Presented method can help decrease those deviations and hold process parameters automatically as needed for the performed task to ensure product quality.

1 Introduction

By analyzing our surroundings we can find from a dozen to hundreds of printed circuit boards placed in a different subjects ranging from gardening tools, mobile phones, and ending on supercomputers. Development trends and especially miniaturization turn into an increase of the requirements for circuit geometry producers have to face. Exposure is one of the important steps in the production process of printed circuit boards. It fulfills two important tasks. Initially is responsible for the very accurate image transfer of the connections from the film to the double-sided copper resist covered laminate, and on the end of the production process appear again to prepare base to electronic assembly by imaging the solder mask. The accuracy of both processes is particularly important in the case of multi-layer printed circuit boards, where subsequent stages of the manufacturing operations leading them to integration into a single circuit. Typical multilayer circuit production process is complex (Fig. 1) and consists a number of successive stages, such as preparation of laminates, resist coating, primary image exposure, developing, etching, realizing automatic optical inspection, bonding, pressing, drilling, exposing outer layers, and coming back to developing, etching, then printing solder masks, applying labels,

Jarosław Zdrojewski · Adam Marchewka

Institute of Telecommunication, University of Technology and Life Sciences, Kaliskiego 7,
85-791 Bydgoszcz

e-mail: {jaz, adimar}@utp.edu.pl

routing and electrically testing [1, 5]. The technology is so complex, many steps in the process can and will bring the deviations, and the final goal is electrical reliability and mechanical stability. A compromise between the density of elements and process capabilities must be maintained to provide desired multilayer circuit precision. Limitations or errors in the manufacturing process will result in mutual displacement of layers and the geometrical shape deviation which will prevent the various internal layers being reliable connected using drilled and metalized holes.

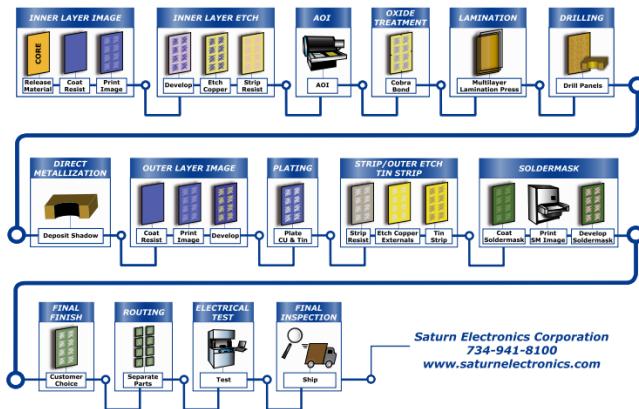


Fig. 1 PCB production cycle presenting localization of the exposure process in the PCB production as one of the first middle and last step in the cycle [11]

The article describes the steps to reduce significantly the position deviation accompanying the exposure process. Considerations will address the possibility of automatic film geometry control, in particular restrictions on the method proposed by the authors. Exposure process of the multilayer circuit board is realized by printing units equipped with exposure frames (Fig. 2a). Regular exposure frame will take two films fixed with vacuum to the upper and lower glass. Such films are relative positioned to each other in the inner layer mode or to the holes in the drilled panel fastened using two pins. Films have marks used to detect (Fig. 2b) multipoint location of the film/panel position. The exposure process based on 2 strong UV sources can produce heat affecting the films by introducing a thermal geometrical deviation that could influence the crucial outcome of imaging process. The objective is to minimize the Thermal impact on the final result can be minimized by precise measurement and control of the associated deviations.

One of the main ideas for this paper is to analyze limitations of previously introduced and presented algorithms [6, 10]. Similar problem of detecting specific objects in digital images was presented in [2]. Because of place where results of image processing is used in the exposure unit control process, crucial is to find out if proposed strategy can significantly impact and improve process stability or is limited, if so then on what level.

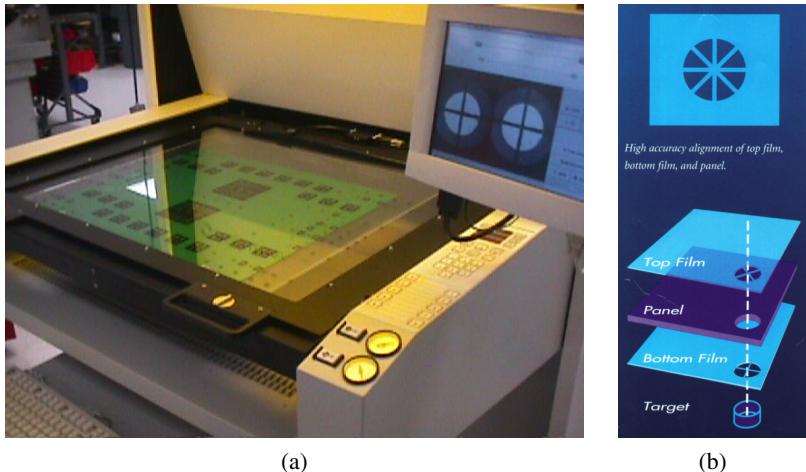


Fig. 2 Single tray of exposure machine equipped with a digital cameras (a), Films and panel targets layout used for relative position verification (b)

2 Targets Analyse

Method proposed in [6, 10] where implemented using the Radon transformation based registration marks edge detection [3, 5, 8, 9]. The main assumption of the method was, to consider analyzed image can be described as a sum of the individual components:

$$f(x, y) = f_l(x, y) + f_o(x, y) + f_e(x, y) \quad (1)$$

where $f_l(x, y)$ - a straight lines component of the image, $f_o(x, y)$ - circles component, and $f_e(x, y)$ - noise.

The proposed algorithm consists of two independent stages:

- lines detection in the image (two edges to determine single marker line),
- circle detection (position for circle axis of symmetry).

Lines detection in the image $f(x, y)$ is limited to the search finding local maxima (peaks) in the image transform. To limit the information contained in the RT coefficients original image is transformed with the use of Sobel filter. Radon transform is applied to the resulting image containing edges information. In the resulting array of RT coefficients, depending on the class of the analyzed image, search of 8 up to 10 local maxima is performed $\max_{local}(R_\theta(x'))_n$ for $n \in \{1, \dots, 10\}$. Using located angle θ and x' values position of individual straight lines can be calculated, depending on the position of individual marks needed to proceed with a alignment task the device.

Ideally in consideration, each target (mark) line is defined by two parallel lines. Which means that among the local maxima found for this purpose, the local maxima

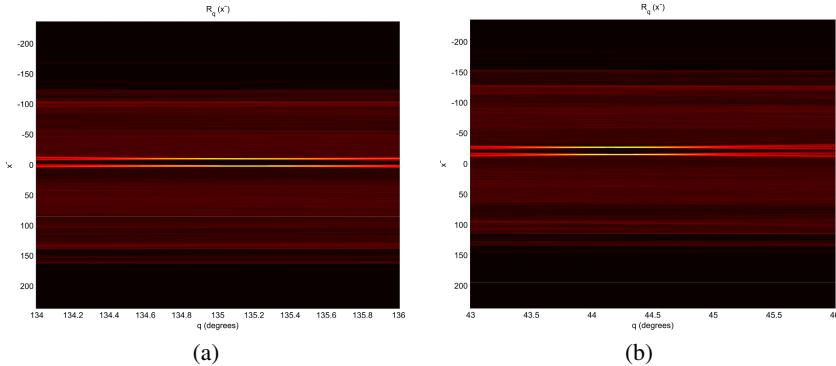


Fig. 3 Radon Transform for the horizontal line edges of the film mark

should be found grouped in pairs, which $R_\theta(x'_n) - R_\theta(x'_m) \in \{0^\circ, 180^\circ\}$ where $n, m \in \{1, \dots, 10\}$.

During the implementation of the method (especialy the particular images) not every angle was considered to be taking in account, but only those for which we expect to detect mark line as follows: $0^\circ, 45^\circ, 90^\circ, 135^\circ$. Of course it was necessary to check a limited range of tolerance for a searched lines inclination. So Radon transform (Fig. 3) is performed only for certain angle ranges θ . Assumed tolerance is $\pm 2^\circ$ and the angle during analyze will vary 0.001° .

With such approach it was possible to increase the method accuracy with maintaining an appropriate, desired respond rate for the device (exposure unit).

Already during the experiments proved that the designated lines indicating edges of the target marks line are not perfectly parallel to each other. Angle in between them can vary within the limits $0 \div 0.001^\circ$. This is due to the imprecision of the image acquisition elements (introduced by optic, lens, prisms and mirrors and add, electronic components like camera sensor, A/D converter) and edge detection algorithm. Taking into consideration the real image resolution is defined by optic at $1\text{px} = 15\mu$ and maximum length of the edge of the pointer is 350px , an error at the 0.35px level or 5.25μ has to be taking in account. This is a significant measurement error, but however, it is important to remember, that edges of the given mark are used only to calculate relative location in between marks located on the single image. Therefore, to represent the mark auxiliary line l_d should be determined as a line parallel to the calculated edges. In most cases this will be the axis of symmetry for analyzed pair of lines. However, in other cases angle bisector l_d should calculated for lined defining mark.

Lets $l_1 : A_1x + B_1y + C_1 = 0$ i $l_2 : A_2x + B_2y + C_2 = 0$ - will represent line for the edges of the single mark. Angle bisector in between them can be marked as fallow:

$$\frac{|A_1x + B_1y + C_1|}{\sqrt{A_1^2 + B_1^2}} = \frac{|A_2x + B_2y + C_2|}{\sqrt{A_2^2 + B_2^2}} \quad (2)$$

We have to solution of the equations (2).

$$l_{d_{1,2}} = \begin{cases} \frac{A_1x+B_1y+C_1}{\sqrt{A_1^2+B_1^2}} = \frac{A_2x+B_2y+C_2}{\sqrt{A_2^2+B_2^2}} \\ \frac{A_1x+B_1y+C_1}{\sqrt{A_1^2+B_1^2}} = \frac{-A_2x-B_2y-C_2}{\sqrt{A_2^2+B_2^2}} \end{cases} \quad (3)$$

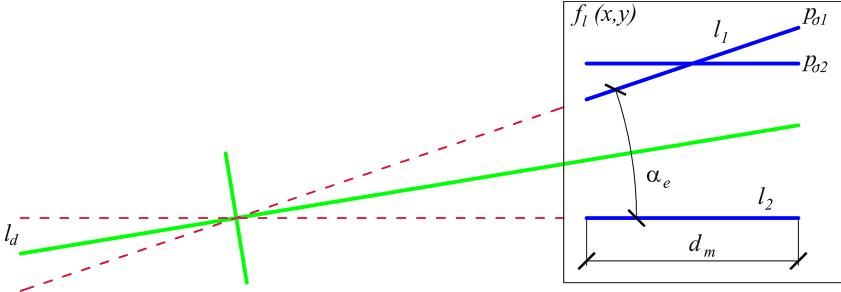


Fig. 4 Calculating intersection point and center line to define mark line location

For each of the four markers auxiliary line l_d must be calculated (Fig. 4). But only for a couple $\{l_{d_1}, l_{d_2}\}$ intersection point should be calculated to define film location in relation to the other film mark or to the hole in the panel depending on the type of process. We assume that mark lines are perpendicular to each other. In such case the solution of the problem is limited to intersection point of two angle bisectors finding. Despite varying amounts of possible solutions there is only one belonging to the field of the image.

For the above analysis, we will define the measurement error. Let's d_m will be the length of the mark line, α_e - maximum angle deviation for two lines limiting the the mark line and defined as a distance between points p_{o1} i p_{o2} satisfying the following conditions:

$$\begin{aligned} p_{o1} &= \begin{cases} y = -x \\ y = -\tan(180 - \alpha_e/2)x + \tan(\alpha_e/2)d_m/2 \end{cases}, \\ p_{o2} &= \begin{cases} y = -x \\ y = -\tan(\alpha_e/2) - \tan(\alpha_e/2)d_m/2 \end{cases}. \end{aligned} \quad (4)$$

Substituting known values, our measurement error is: $0.015\text{px} \cdot 15\mu\text{m} = 0.225\mu\text{m}$.

3 Experimental Results

Series of tests for real images and sample panels and films registered by the exposure machine were conducted. The group contained images whose double sided registration panel. This type of processes include both top, and bottom film and the

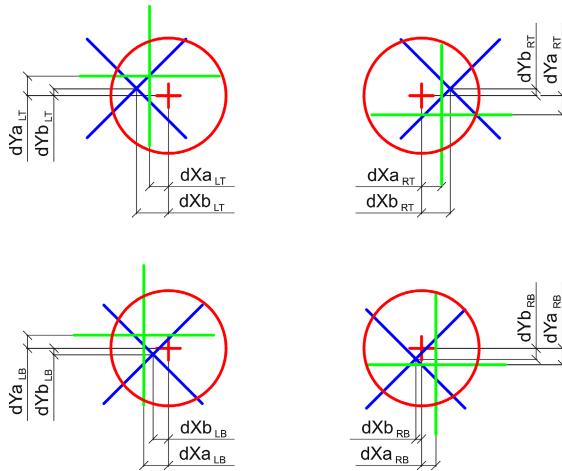


Fig. 5 Dimensional deviation and alignment deviation [dX, dY] alignment deviations (LT Left top camera, RT right top camera, LB Left bottom camera, RB right bottom camera), PD process deviation

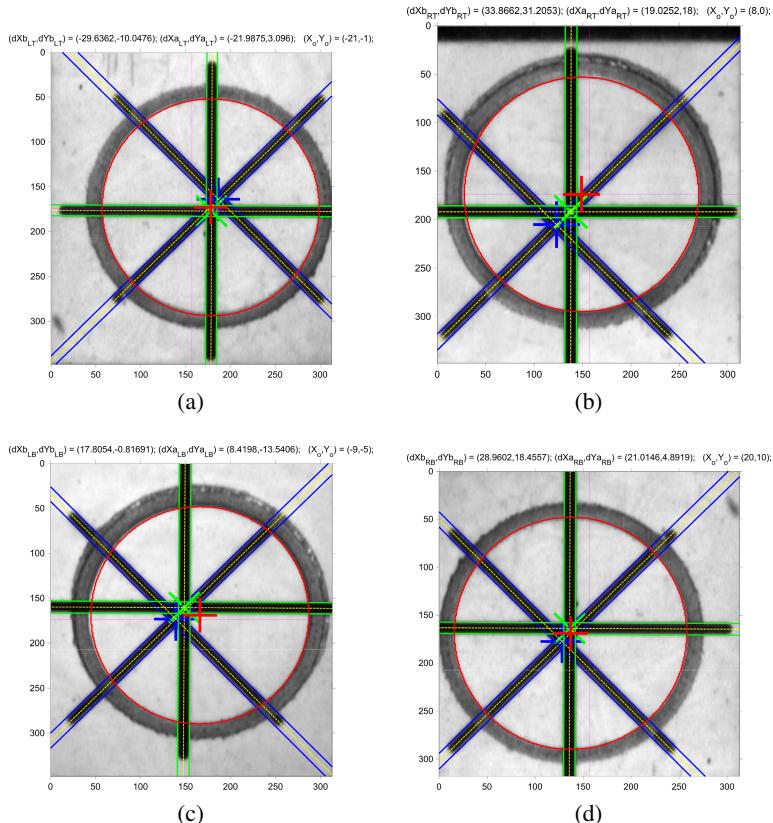


Fig. 6 Images with detected marks for four cameras system

panel targets (Fig. 5). For this group trials were conducted to calculate values defining films deviations (Fig. 6). All deviation are calculated in the way taking in account difference between level of deviation (single micrometers) and distance between targets (usually hundreds of millimeters). In this case ignoring the dY component when calculating dX deviation and ignoring dX when calculating dY will not create error bigger than 0.05μ .

4 Conclusion

This article presents an approach to analyze of automatic films dimensions control using new optimized method. Goal was to analyze limitations of previously introduced and presented algorithms. Results of image processing are used in the exposure unit control process. All measurements are important in the proposed strategy and can significantly impact and improve exposure process stability. Films in exposure unit running on the base of standard imaging technology are crucial for the process fidelity. Any thermal or initial based on laser plotter error mistake will be copied during the exposure cycle over several or hundreds of panels, finally leading to producing panels for scrap. All film measures data is verified using the predefined tolerance levels and finally resulting information is saved in the database for the process optimization purpose. The proposed method of getting inputs from marks on the films is based on Radon Transform. Beside image analyze, research process was extended by following film and panel geometry verification step and calculating deviations having significant impact on the production process. Proposed algorithm can detect marks fast and correctly. Because of its μ resolution it allows to eliminate mistakes of image processing methods' imperfections. The presented method was tested with film marks with a low modulation grade and blurred symbol images. It is necessary to remember that stability of the process can be illumination dependent especially in the production conditions of a yellow room. In some cases especially for panels with resist covered holes top film mark image is deformed. However, in the future work an analysis of border conditions to detect marks and its limitations will be considered.

References

1. Coombs Jr., C.F.: Printed Circuit Handbook, 6th edn. McGraw-Hill (2007)
2. Forczmański, P., Frejlichowski, D.: Robust stamps detection and classification by means of general shape analysis. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part I. LNCS, vol. 6374, pp. 360–367. Springer, Heidelberg (2010)
3. Gur, E., Weizman, Y., Zalevsky, Z.: Radon transform based image enhancement for microelectronic chips inspection. IEEE Transactions on Device and Materials Reliability (2010)

4. Holden, H.: Calculate Your Fabrication Capability Coefficient, Happy Thoughts, CircuiTree, BNP, pp. 23.5 (February 15, 2006)
5. Leavers, V., Boyce, F., The, J.F.: radon transform and its application to shape parametrization in machine vision. *Image Vision Comput.* 5, 161–166 (1987)
6. Marchewka, A., Zdrojewski, J.: Film Alignment before Solder Mask Exposure. In: Choraś, R.S. (ed.) *Image Processing and Communications Challenges 3*. AISC, vol. 102, pp. 387–393. Springer, Heidelberg (2011)
7. Ohlig, B.: Registration and Knowledge Through and Across Your Panel, CircuiTree (April 1, 2002)
8. Radon, J.: On the determination of functions from their integrals along certain manifolds. In: The radon transform and some of its applications, annexe a, Wiley, New York (1983); translation of Radon's 1917 paper by R. Lohner (1917)
9. Toft, P.: Detection of lines with wiggles using the radon transform (1996)
10. Wocianiec, R., Zdrojewski, J.: Film Position Optimization Before Exposure. *Image Processing & Communication An International Journal* 5(2), 35–46 (2010)
11. <http://www.free-press-release.com/news-visit-pcb-manufacturing-saturn-electronics-corporation-in-booth-1628-at-ipc-apex-2010-in-las-vegas-1270505048.html>

Structure from Motion in Three – Dimensional Modeling of Human Head

Anna Wójcicka and Zygmunt Wróbel

Abstract. A human face is an important component of the model form. Is the main instrument of communication and expression of emotions and personality so three-dimensional model of the head must imitate exactly the actual anatomy and features which is one of the most difficult tasks facing graphic. The paper presents a model of a human head reconstruction using structure from motion - the acquisition of three-dimensional models of a series of 2D images. In detail describes the mathematical basis of the method and then refers them to build the model in a computer program. It has been shown that due to the very good quality three - dimensional model, the method of Structure from motion is a competition in modeling objects for standard graphic programs.

1 Introduction

The creation of computer graphics led to a revolution in the world of visual effects and film. As far as evolution 3D modeling, analysis and processing of images, animation and speech synthesis become more advanced, hardware faster and cheaper, computer graphics spread on the internet, games, commercials and movies. The first feature film which shown in fully animated form was ‘Young Sherlock Holmes’ (1986). The bigger challenge was to use computer graphics to create realistic models of creatures and people. ‘Jurassic Park’ (1993) is the first example of a film which

Anna Wójcicka

Pedagogical University of Cracow, Institute of Technology, 2 Podchorazych Str.,
30-084 Cracow, Poland

e-mail: aniawojcicka@gmail.com

Anna Wójcicka · Zygmunt Wróbel

University of Silesia, Institute of Computer Science,
Department of Computer Biomedical Systems, 39 Bedzińska Str., 41-200 Sosnowiec
e-mail: zygmunt.wrobel@us.edu.pl

fully integrated computer form the live action. The dream of computer graphics was to create a fully synthetic human form indistinguishable from a real person, examples may be movies: ‘Tony de Peltre’ (1985), ‘Rendez-vous a Montreal’ (1988), ‘Sextone for President’ (1989). But it was obvious to the audience that they are computer-generated. Jar-Jar Binks from movie ‘Star Wars: Episode I’ (1999) is the first fully articulated synthetic humanoid actor. Recent advances in lighting modeling and texturing led to create more realistic forms. The ‘Final Fantasy: The Spirits Within’ (2001) movie is described as the most ambitious attempt to use synthetic actors to portray in the film realistic people [18].

A human face is an important component of the model form. Is the main instrument of communication and expression of emotions and personality. Realistic animation is one of the most difficult tasks facing for animator and graphic. Model face must imitate exactly anatomy and features real. It should allow the geometric representation of a person. In the literature this is called compatibility [19] or statistics [22]. Face models are used in the production of movies, cartoons, in animation and realistic in telecommunications and human-computer interaction (HCI). In the cartoon, face model should give personnel mines and personalities, often exaggerated. In realistic animation, face shape and movements must be compatible with human anatomy. In applications to communications, human-computer interaction interact and there is an additional requirement of computational efficiency. In some applications, the model’s face must exactly match a person.

The paper presents a model of a human head reconstructed using Structure from motion (SfM). The head model is reconstructed from three-dimensional models build based on a series of 2D images. The mathematical fundamentals of the SfM have been described in details referred to the reconstruction of the 3D model using Agisoft PhotoScan.

2 Methods

Nowadays, create model of a human head from scratch was no easy task. Transfer of static properties of the model geometry in a computer program is a challenge. Standard face models can be constructed as a flat polygon mesh or the surface of a higher degree. Mostly due to the simplicity, are use the multi-angle models [19, 26]- relatively easy to set up and modeling . There are three methods for interactive design applied in the creation of models: wire method (low resolution), modify the simple shape represented with high resolution and design the area of a high-resolution [18]. These methods for creating three-dimensional models are often not easy to make because they require a lot of experience and patience. The more realistic models need more time to devote to his execution.

Structure from motion (SfM), developed in the 80’ [24] is a technique for obtaining information about the geometry of 3D scenes from 2D images [18]. Due to the low computational capabilities of computers at that time, it had no followers. Currently, the computer programs based on SfM gaining a large popularity, as an

easy way to allow perform three-dimensional model of an object. It is studied in the fields of computer vision and visual perception, which has a number of applications in different areas, such as archeology [20], topography [11], psychology [25]. An interesting example is the use of SfM in Google street application [5]. Structure from motion techniques are used in a wide range of applications including photogrammetric survey [12], the automatic reconstruction of virtual reality models from video sequences [28], and for the determination of camera motion footage of real-world scenes [4, 13].

The most important challenges aims to recover 3D models of usually rigid objects from an uncalibrated or calibrated sequence of 2D images. SfM method to create a 3D model can be divided into two main phases: retrieving camera motion and reconstructing a 3D model. The first stage of - retrieving camera - consists of two steps: estimating the fundamental matrix [2, 8], obtain the motion of the camera. The second stage - reconstructing a 3D model - also consists of two steps which are the recovery of 3D scene points and rendering of a 3D model As it presented in figure 1. [4].

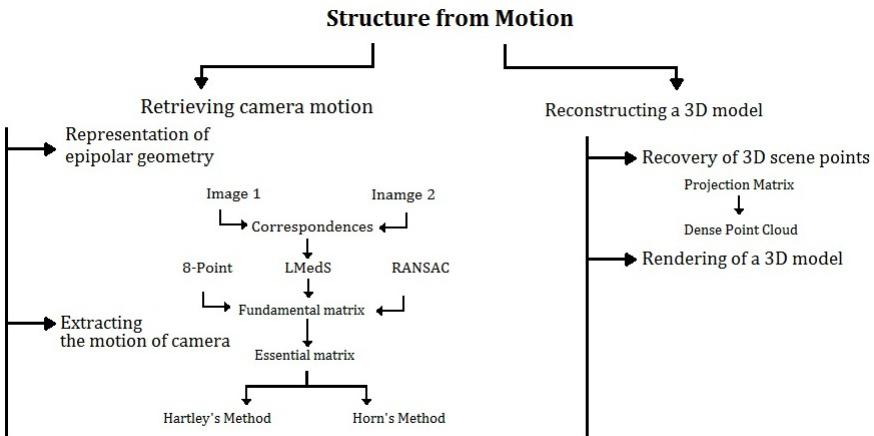


Fig. 1 Action scheme of Structure from Motion method

The first stage in the creation of 3D models from 2D images is searching for corresponding points in the photos, adjusting them, then determine the position of the camera for each shot and refines camera calibration parameters. The result is the sparse point cloud and a set of camera positions are formed. The point cloud represents the results of photos alignment [17]. In the next stage of the received point cloud model is reconstructed in 3D and rendered. The basis of this method is the epipolar geometry, which describes the relationship between position of camera systems one to the other, presented in figure 2 – define corresponding points in an image pair. This relation called the fundamental matrix or the essential matrix, and it is represented by a 3×3 singular matrix.

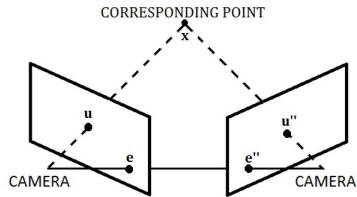


Fig. 2 The epipolar geometry. Scene 3D point X is projected to the two images giving raise to a correspondence (u, u'') . The camera projection centers Camera and Camera'' and the 3D point X form an epipolar plane which constrains the positions of u and u'' to the corresponding epipolar lines. Epipolar lines of different scene points intersect in the epipoles e and e'' in the first and second image respectively

Essential matrix describes the pair of calibrated views defined in intrinsic camera parameters. Intrinsic camera parameters can be obtained by calibration which is performed before or after the taking of images. Alternatively it can be computed from a self-calibration within the process of SfM. The Essential matrix was invented by Longuet-Higgins in [16]. Longuet-Higgins describes [16] how an essential matrix relating a pair of calibrated views can be estimated from eight or more point correspondences by solving a linear equation [4]. If cameras have been uncalibrated, camera motion is calculated from fundamental matrix. There are some methods for estimating fundamental matrices presented. These are the 8-Point Algorithm [16], the Least Median of Squares (LMedS) Algorithm [21], and the Random Sample Consensus RANSAC Algorithm [3]. Essential is a specific matrix will fall of fundamental matrix, in which the image parameters are normalized. The equation $E = k_2^T F k_1$ presented relationship between the essential matrix E and the fundamental matrix F . Assume matrices k_1 and k_2 represent the matrices of intrinsic camera parameters.

Next step is extracting the motion of the camera. In 2D images are recorded the camera motion information (the orientational and positional of camera). This information are necessary to create 3D model. Differences in camera motion resulting from the method of obtaining images that come from a single camera movement when shooting or in case two moving cameras. The motion can be represented by a 3×3 rotation matrix and a 3×1 vector. Camera motion can also be acquired from the fundamental matrix from at least seven point correspondences or extracted from splitting the essential matrix. For computing camera motion from essential matrix can be use methods which were proposed by Horn [7], and by Hartley and Zisserman [10].

Next step is extracting the motion of the camera. In 2D images are recorded the camera motion information (the orientational and positional of camera). This information are necessary to create 3D model. Differences in camera motion resulting from the method of obtaining images that come from a single camera movement when shooting or in case two moving cameras. The motion can be represented by a 3×3 rotation matrix and a 3×1 vector. Camera motion can also be acquired from the fundamental matrix from at least seven point correspondences or extracted from

splitting the essential matrix. For computing camera motion from essential matrix can be use methods which were proposed by Horn [7], and by Hartley and Zisserman [4, 10, 18]. The obtained parameters of camera motion and corresponding points extracted from pairs of images are required for the reconstruction of the 3D model. This stage of SfM aims to create a point cloud consisting of relative positions of all point correspondences. In the next stage of the cloud of points is subject to the rendering process. All points belonging to the cloud are defined separately in the space, do not contain information on local connectivity between points. Triangulation method local point cloud points are the forms of a triangular mesh surface. Model surface is divided into triangles in such a way that each wheel described in the three neighboring points does not contain any other point. It is Delauney triangulation [1], which is widely used in the visualization. It is obvious that the 2D triangulation is more effective than the 3d. Therefore, prior to triangulate the 3D model, the cloud point was projected into one of the images. Triangulation is performed on this image - projection. Then point clouds are combined based on the triangulation of its projection. Based on the triangular mesh we can render a surface using a graphics rendering library, for example OpenGL [23].

3 Experiment

Nowadays, many programs can perform 3D model by Structure from motion for example: Agisoft PhotoScan, 123D Catch or Arc3D. For the purpose of this work AgiSoft PhotoScan has been selected due to easy user interface and advanced mode of operation. Pictures can be taken with any standard digital camera with reasonably high resolution (5 megapixel or more). Wide-angle lenses be better adapted to mapping relationships between objects in space than telephoto lenses. Fish eyes and ultrawide angle lenses are poorly modeled by the distortion model implemented, which leads to inaccurate reconstructions. Avoid on the photos not textured, shiny and transparent objects, flat surfaces. To perform 3D object model glossy photo must be carried out under a cloudy sky, which causes the least amount of reflections. In the pictures should appear moving objects and unwanted foregrounds. Best results are obtained when the information is captured from multiple viewpoints – three or more, with large number of images. Unfortunately, the number of images entails hardware limitations. None of the pictures do not cut and converted into geometrically, because this images will be a source of errors and incorrect results in reconstructed model. Photometric modifications do not affect reconstruction results. Calibrated cameras have the information about camera motion that can be use to estimate the field of view of the photo. It is save in EXIF part of photo. If the data are available, you can expect the best results during the reconstruction model. Otherwise, the program automatically assumes that the 35mm focal length equivalent equals to 50 mm and tries to align the photos in accordance with this assumption. If the focal length is significantly different from 50 mm the alignment can give incorrect results or even fail. In this case, it is necessary to manually calibrate the camera [6, 17].

3.1 Aligning Photos

Three – dimensional model of a human head has been reconstructed from a series of 44 calibrated images, made from three viewpoint. Pictures were taken with a standard digital camera with wide angle lens. Schematic of series of photo, made in different viewpoint is shown in figure 3. Then, loaded images into computer program - Agisoft PhotoScan. Then the images were subjected to Aligning photos. As a result of this process, the program detects camera motion and then on the basis of individual images sets the corresponding point – figure 3.

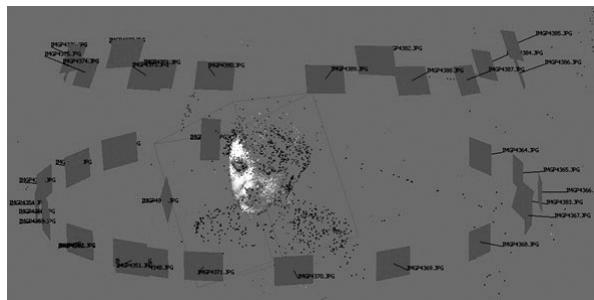


Fig. 3 Dialog box from Agisoft PhotoScan - aligning photos, which show diagram of visualization of the images position in space. These images are aligned by the program in order to prepare the data for the reconstruction of the three - dimensional model.

As can see, it detected motion camera viewpoint in which the images were made. Alignment parameters can be modified in the Align Photos dialog box. It is improve the process of detecting the camera motion is recommended to set higher accuracy which Increases processing time. In the initial stage, can be use a lower accuracy. Aligning photos process with a large number of photos may take a relatively long time, which is intended mainly on matching of detected features across the photos. This time can be accelerated by applying photo mask elements which can't be reconstructed in a 3D model. Mask used to reconstruct a model of a human head. At this stage, it can still manually adjust the wrong position of photos, if any are detected. At the moment a cloud of points that can be exported to other graphics program that will allow the construction of 3D model.

3.2 Building Model Geometry

Next step is building model geometry. It can choosen several methods of reconstruction. It is recommended to start to build a model with the lowest quality, evaluate the results and then recompute the results using a higher quality setting. As shown

in figure 4, the human head was separated from the background using a mask, so part of the background submissive inaccurate reconstruction. All of these errors can be removed during the selection mask (but it should be done for all the images separately) or on the finished model after exporting the graphics program. The triangulation model shown in figure 5.



Fig. 4 Geometrical model of the human head



Fig. 5 The triangulation model of human head

3.3 Building Model Texture

Building model texture is the next step of building three - dimensional model of human head. Several methods are available texture mapping model in the Build dialog box Texture: generic-default mode, adaptive orthophoto-applied to flat surfaces such as buildings, orthophoto, single photo, keep uv. The result of facial skin texture mapping shown in figure 6.

Ready 3D model can be saved in various formats to allow further work in a graphics program such as Wavefront OBJ, 3DS file format, VRML, COLLADA, Stanford PLY, Autodesk DXF, U3D, Adobe PDF exporting results. In addition there are useful function which enables Agisoft PhotoScan software export the data to the cloud geometry and and texture as separate files.



Fig. 6 Facial skin texture mapping: a) anterior view b) anterior - right view (magnification of upper face) c) left view d) anterior-top view

4 Conclusions and Future Work

The reconstruction of the human head performed by SfM method has been successfully accomplished. As it can be seen in the results the presented method is efficient to create a model of a human head. Quality of three-dimensional image with textures gives the photorealistic effect which suggests spatiality of image. The quantitative measurement of the image quality was considered as a more advanced and extended task designated to describe in a separate paper. Similar to the reported in the literature and discussed in the introduction achievements in the standard modeling, the results obtained with the SfM does not deviate from the well-known.

Structure from motion opens up the possibility of modeling difficult to obtain 3D models without use of standard methods for drawing graphics programs. A major impact on the quality of 3D model is accurate shots of images and their quantity. Due to high accuracy the mesh, it can modify 3D model in graphics programs (Maya, 3D StudioMax).

In our future work we will focus on improving method effectiveness and efficiency which may be a strong side of SfM and the software. This parameters should be the subject of detailed research and metrology, depending on the resolution, number of images and the hardware on which the reconstruction takes place.

Acknowledgements. Anna Wójcicka is beneficent of “Silesian Cooperation: Innovations For Efficient Development (SWIDER)“ project realized within Human Capital Operational

Programme, Priority VIII: Regional human resources for the economy, 8.2 Transfer of knowledge, Sub-measure 8.2.1 Support to cooperation of scientific environment and enterprises. Project partially funded from European Union Project based on European Social Funds.

References

1. Delaunay, B.: Sur la sphére vide, *Izvestia Akademii Nauk SSSR. Otdelenie Matematicheskikh i Estestvennykh Nauk* 7, 793–800 (1934)
2. Faugeras, O.: What can be seen in three dimensions with an uncalibrated stereo rig? In: Sandini, G. (ed.) *ECCV 1992. LNCS*, vol. 588, pp. 563–578. Springer, Heidelberg (1992)
3. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM* 24, 381–385 (1981)
4. Liu, G.: Structure from Motion — A Case Study for Reconstructing 3D Objects. The University of Auckland CITR, Tamaki Campus (2006)
5. Google Street View Pittsburgh Research Data Set, Provided and copyrighted by Google
6. Havlena, M.: Incremental Structure from Motion for Large Ordered and Unordered Sets of Images. PhD Dissertation, Center for Machine Perception Department of Cybernetics Faculty of Electrical Engineering Czech Technical University in Prague (2012)
7. Horn, B.: Recovering baseline and orientation from essential matrix, <http://people.csail.mit.edu/bkph/articles/Essential.pdf>
8. Hartley, R.: Estimation of relative camera positions for uncalibrated cameras. In: Sandini, G. (ed.) *ECCV 1992. LNCS*, vol. 588, pp. 579–587. Springer, Heidelberg (1992)
9. Hartley, R., Sturm, E.: Triangulation. In: Proc. ARPA IUW, Monterey, California, pp. 957–966 (1994)
10. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2000)
11. Irschka, A., Zach, C.: Bischof H Towards wiki-based dense city modeling. In: *Virtual Representations and Modeling of Large-scale environments (VRML)*, pp. 1–8 (2007)
12. Kraus, K.: *Photogrammetry*, vol. I & II. Dummler (1997)
13. Kutulakos, K., Vallino, J.: Calibration-free augmented reality. *IEEE Transactions on Visualization and Computer Graphics* 4(1), 1–20 (1998)
14. Koprowski, R., Wrobel, Z.: The cell structures segmentation Computer Recognition Systems. In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierk, A. (eds.) *Computer Recognition Systems. ASC*, vol. 30, pp. 569–576. Springer, Heidelberg (2005)
15. Koprowski, R., Wrobel, Z.: Layers recognition in tomographic eye image based on random contour analysis. In: Kurzynski, M., Wozniak, M. (eds.) *Computer Recognition Systems 3. AISC*, vol. 57, pp. 471–478. Springer, Heidelberg (2009)
16. Longuet-Higgins, H.: Computer algorithm for reconstructing a scene from two projections. *Nature* 293, 133–135 (1981)
17. Manual Agisoft PhotoScan User Manual: Standard Edition, Version 0.9.0, Agisoft LLC (2012)
18. Parent, R.: *Animacja komputerowa, Algorytmy i Techniki*, PWN, Warszawa (2012)
19. Parke, F.: A Parametric Model for Human Faces. Ph.D dissertation, University of Utah (1974)
20. Pollefey, M.: 3D recording for archaeological fieldwork. *IEEE Computer Graphics and Applications* 23(3), 20–27 (2009)

21. Rousseeuw, P., M'Leroy, A.: *A Roboust Regression and Outlier Detection*. Wiley, New York (1987)
22. Rydfalk, M.: CANDIDE: A Parameterized Face”, Technical Report LiTH-ISY-I-0866, Linkoping University, Sweden (1987)
23. Shreiner, D.: OpenGL Reference Manual: The Official Reference Document to OpenGL. Addison-Wesley Professional, Boston (2004)
24. Ullman, S.: Computational studies in the Interpretation of Structure and Motion: Summary and Extension, vol. 706. Institute of Technology; Artificial Intelligence laboratory, A. I. Memo (1983)
25. Vaina, L., Lemay, M., Bienfanga, D., Choi, A., Nakayamaa, K.: Intact “biological motion” and “structure from motion” perception in a patient with impaired motion mechanisms: A case study. *Visual Neuroscience* 5(4), 353–369 (1990)
26. Waters, K.: A Physical Model of Facial Tissue and Muscle Articulation Derived from Computer Tomography Data. In: SPIE Visualization in Biomedical Computing, pp. 574–583 (1808, 1992)
27. Zhang, Z.: Determining the epipolar geometry and its uncertainty: a review. *Int. J. Computer Vision* 27, 161–195 (1998)
28. Zisserman, A., Fitzgibbon, A., Cross, G.: VHS to VRML: 3D graphical models from video sequences. In: ICMCS, pp. 51–57 (1999)

A Short Overview of Feature Extractors for Knuckle Biometrics

Michał Chorąś

Abstract. In this paper an overview of image processing methods for feature extraction applied to knuckle biometrics also termed as FKP (finger-knuckle-print) is presented. Knuckle is a part of hand, and therefore, is easily accessible, invariant to emotions and other behavioral aspects (e.g. tiredness) and most importantly is rich in texture features which usually are very distinctive. In this paper a short overview of the known recent approaches to human identification on the basis of knuckle images is given.

1 Introduction

Even though biometric identification systems became our reality and are no longer science-fiction visions, only several modalities have been widely deployed and such systems still have many drawbacks. The most known and often used modalities are fingerprints, face, hand geometry and iris. For example fingerprint and faces are widely deployed in large-scale systems such as border control, asylum control and biometric documents. But due to the problems with large-scale scalability, security, effectiveness and last but not least user-friendliness and social acceptance new emerging modalities are still investigated. One of such new and promising modalities is human knuckle.

In this paper the overview of feature extractors for such emerging modality, namely knuckle or finger-knuckle-print (FKP) is presented. Knuckle is a part of hand, and therefore, is easily accessible, invariant to emotions and other behavioral aspects (e.g. tiredness) and most importantly is rich in texture features which usually are very distinctive. Knuckle biometrics methods can be used in biometric systems for user-centric, contactless and unrestricted access control e.g. for medium-security access

Michał Chorąś

Image Processing Group, Institute of Telecommunications, UT&LS Bydgoszcz
e-mail: chorasm@utp.edu.pl

control or verification systems dedicated for mobile devices (e.g. smartphones and mobile telecommunication services).

The sample knuckle image from IIT Delhi Database is presented in Figure 1 [1].

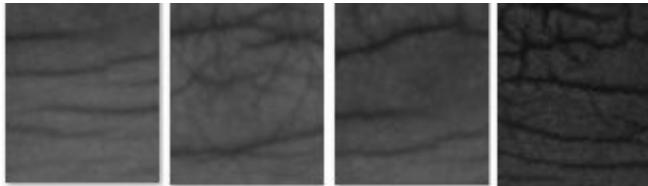


Fig. 1 Sample knuckle images from IIT Delhi Database [1]

2 Feature Extractors for Knuckle Biometrics

Most of the applied methods originate from known signal processing transformations or image processing methodologies. In general those can be categorized as approaches based on:

- Gabor-based approach (including e.g. 1D Log-Gabor),
- Ridgelets and transforms (Radon, Riesz),
- (Probabilistic) Hough transform (PHT),
- SIFT and SURF,
- Phase correlation functions.

2.1 Gabor-Based Features

It seems that Gabor based feature extraction is most common approach for knuckle biometrics. Of course in such approach (due to high dimensionality of Gabor filtering output) various dimensionality reduction techniques are also engaged.

The general function of the two-dimensional Gabor filter family can be represented as a Gaussian function modulated by a complex sinusoidal signal [2]. Specifically, a two-dimensional Gabor filter $\psi(x, y; \sigma, \lambda, \theta_k)$ can be formulated as:

$$\psi(x, y; \sigma, \lambda, \theta_k) = \exp\left(-\frac{x_{\theta_k}^2 + \gamma^2 y_{\theta_k}^2}{2\sigma^2}\right) \exp\left(\frac{2\pi x_{\theta_k}}{\lambda} i\right), \quad (1)$$

where $x_{\theta_k} = x \cos \theta_k + y \sin \theta_k$; $y_{\theta_k} = -x \sin \theta_k + y \cos \theta_k$; σ is the standard deviation of the Gaussian envelope along the x - and y -dimensions; γ is the spatial aspect ratio and; λ and θ_k are the wavelength and orientation, respectively [3] [4].

For example, Shariatmadar et al. [5] apply Gabor feature extraction followed by PCA and LDA for four fingers. Then feature level fusion is made before the match-

ing process. Yang et al. [6] use Gabor filtering features and OLDA (Orthogonal LDA). Similarly, 8 Gabor filtering orientations and 5 scales are also used in FKP recognition based on Local Gabor Binary Patterns ([7]).

Meraoumia et al. [8] propose to use result of 1D Log-Gabor filtering in a palm-print and knuckle multimodal system. 1D Log-Gabor filters were also used by Cheng et al. [9] to calculate features of knuckle images in a contactless scenario (images acquired by smartphones).

Zhang et al. ([10]) proposed to use Gabor filters create so called competitive coding (CompCode) representations of knuckle images. Later they (Zhang et al. [11]) proposed Gabor filtering to create improved competitive coding (ImCompCode) and magnitude coding (MagCode) representations of knuckles. The same authors also proposed to use known Gabor characteristics to obtain ensemble of local and global features/information. They proposed an LGIC (local-global information combination) scheme basing on CompCode as well as Phase Only Correlation (POC) and Band Limited Phase Only Correlation [12].

It is worth to mention that global subspace methods such as PCA, LDA and ICA can be used not only for dimensionality reduction (of. e.g. Gabor filtration based vectors) but also as global appearance feature extractors as shown in [13].

2.2 (*Probabilistic*) Hough Transform

In [14] image is analyzed by means of Probabilistic Hough Transform (PHT), which is used both for determining the dominant orientation and also for building the intermediate feature vector. This feature vector (describing the knuckle texture) is built using the PHT output information, which contains set of line descriptors represented by formula 2, where $LD_i(N)$ stands for $N - th$ line descriptor of $i - th$ image, (b_x, b_y) the Cartesian coordinates of line starting point, (e_x, e_y) the Cartesian coordinates of line end point, θ the angle between the line normal and the X-axis, and d the particular line length expressed in pixels.

The number of extracted lines (N) depends strictly on knuckle spatial properties and varies, therefore these are not directly used to build feature vector.

$$LD_i(N) = [b_{xN}, b_{yN}, e_{xN}, e_{yN}, \theta_N, d_N] \quad (2)$$

Due to the fact the particular knuckle may be rotated, the dominant orientation based on Hough transform is extracted using the θ angle from the line descriptors, which is used to rotate analyzed image in opposite direction to align the dominant line perpendicular to y-axis. After that the y position of particular line and its length is used to build the feature vector. The 30-bins 1D histogram is calculated since the longest and characteristic lines of knuckles are concentrated around one rotation angle.

The vectors obtained by PHT were named "basic" since these are relatively short (one row vector of length 30) and are used for general data set clustering to decrease the number of computations and comparisons of complex features vector in

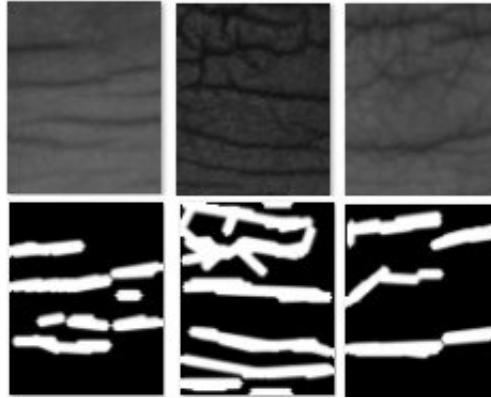


Fig. 2 Sample knuckle images and their representation after applying PHT transform

further phases of our human identification system. The set of line descriptors (eq. 2) obtained from Hough transform are then converted to image representation giving input for matching algorithm.

2.3 SIFT and SURF Feature Extractors

The SIFT stands for scale invariant feature transform. It is often used to extract salient points and is used in many applications such as biometrics or retrieval [15, 16]. SIFT features and their variations are based on image gradients of the pixels in a window around a salient point, and do not take color information into account. Therefore it was proposed as knuckle feature extractor in [17] and [18]. Authors used SIFT transform after enhancing the image, e.g. by Gabor filters.

The SURF stands for Speeded Up Robust Features and is robust image detector and descriptor. It was firstly presented by Herbert Bay in 2006 [19]. It is widely used in object recognition and 3D reconstruction. The key-point of the SURF detector is the determinant of the Hessian matrix, which is the matrix (eq. 3) of partial derivatives of the luminance function.

$$\nabla^2 f(x, y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (3)$$

$$\det(\nabla^2 f(x, y)) = \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 \quad (4)$$

The value of the determinant (eq. 4) is used to classify the maxima or minima of the luminance function (second order derivative test). In the case of SURF the partial derivatives are calculated by convolution with the second order scale normalized

Gaussian kernel. To make the convolution operation more efficient the Haar-like function are used to represent the derivates.

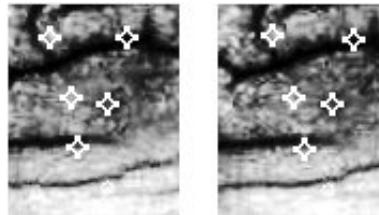


Fig. 3 Detected fiducial SURF points for queering image and its corresponding matches for the template image

If the determinant value is greater than threshold (estimated during experiments on learning data set) then it is considered as a fiducial point. The greater the threshold is the less points (but strong ones) are detected. For each of the fiducial points the texture descriptor is calculated.

In [14] SURF points are used to find the closest matching (if any) between querying image and the templates selected by PHT-based classifier (used by authors in previous step to create intermediate vector (see section 2.2).

Firstly, the points yielding the Hessian determinant value greater than threshold are selected for both querying and the template images resulting in two points data set. Basing on texture descriptors the matching pairs between those sets are found and the outliers (points in one data set that do not have representative in the second data set) are removed. Then the matching cost between those sets is estimated using eq. 5:

$$m_{cost} = \sum_{i=0}^N d(p_i - \frac{1}{N} \sum_{j=0}^N p_j, q_i - \frac{1}{N} \sum_{j=0}^N q_j), \quad (5)$$

where N , d , p and q represents the number of matching pairs, Euclidean distance, point from template image and point from query image respectively. Example of such a mapping is shown in Fig. 3.

Authors of [20] fused both extractors: SIFT and SURF, which allows to describe local patterns (of texture) around key characteristic points. Of course, the images are to be enhanced before feature extraction step.

2.4 Ridgelets and Transforms (Radon, Riesz)

Another approach to knuckle feature extraction is to use ridgelets or known signal processing transforms such as Radon Transform and Riesz Transform. Ridgelets

are able to represent features of images with many lines such as knuckles. Ridgelets are based on Radon transform, in particular on FRAT (Finite Radon Transform). Wavelet transform is used to each projection of the Radon transform. Digital form of ridgelets applied on FRAT is called Finite Ridgelet Transform (FRIT).

Such approach was used in several papers by Goh et al., e.g. in both [21] and [22] in their proposition of bi-modal knuckle-palm biometric system (however, they proposed other feature extractors to palmprints).

In [23] Radon Transform was used, but in different manner. Authors proposed to apply Localized Radon Transform (LRT) for a discrete image to create so called KnuckleCodes which are later matched (similarly to FingerCodes and IrisCodes used for other modalities).

In a recent paper Zhang and Li ([24]) proposed to use Riesz transform to encode local characteristics of knuckle images. Authors successfully applied 1st order and 2nd order Riesz transforms to calculate so called RCode1 and RCode2 and reported a very promising results. Both RCode1 and RCode2 were applied by authors to knuckle and palmprint images.

2.5 Phase Correlation Based Knuckle Similarity Matching

Another approach to represent local knuckle features is based on Phase Correlation Function (PCF) also termed as Phase Only Correlation (POC) [25] [26]. Phase correlation approach relies on Discrete Fourier Transform (DFT) and Inverse DFT and shows similarity between two transformed images.

In most realizations, in order to eliminate meaningless high frequency components (in classic PCF/POC all frequencies are involved), the Band-Limited Phase Only Correlation (BLPOC) is used [25]. In [27] PCF was used for both knuckle and palmprint in a bi-modal system.

3 Conclusions

Knuckle of finger-knuckle-print modality becomes an emerging trend in biometrics and computer vision communities. The major goal of this paper was to briefly present image processing based feature extractors applied to the task of human identification on the basis of knuckles. It was not author goal to compare, assess or evaluate the described approaches.

It is worth to note that most proposed methods and papers are rather recent which proves growing interest in knuckle biometrics.

Apart from applying various, mostly texture oriented methods, the current noticeable trends are also:

1. application of knuckles in multi-modal biometric systems, mainly with palmprint, hand features or hand veins [8] [21] [24] [28],

2. application of knuckles in contactless (touchless) scenarios, especially with images acquired by mobile phones [9] [21] [29].

References

1. http://webold.iitd.ac.in/biometrics/knuckle/iitd_knuckle.htm
2. Gabor, D.: Theory of communication. *Journal of the Institute of Electrical Engineers* 93, 429–457 (1946)
3. Daugman, J.G.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* 2, 1160–1169 (1985)
4. Fogel, I., Sagi, D.: Gabor filters as texture discriminator. *Biol. Cybernet.* 61, 103–113 (1989)
5. Shariatmadar, Z.S., Faez, K.: A Novel Approach for Finger-Knuckle Print Recognition Based on Gabor Feature Extraction. In: Proc. of 4th International Congress on Image and Signal Processing, pp. 1480–1484 (2011)
6. Yang, W., Sun, C., Sun, Z.: Finger-Knuckle Print Recognition Using Gabor Feature and OLDA. In: Proc. of 30th Chinese Control Conference, Yantai, China, pp. 2975–2978 (2011)
7. Xiong, M., Yang, W., Sun, C.: Finger-Knuckle-Print Recognition Using LGBP. In: Liu, D., Zhang, H., Polycarpou, M., Alippi, C., He, H. (eds.) ISNN 2011, Part II. LNCS, vol. 6676, pp. 270–277. Springer, Heidelberg (2011)
8. Meraoumia, A., Chitroub, S., Bouridane, A.: Palmprint and Finger Knuckle Print for efficient person recognition based on Log-Gabor filter response. *Analog Integr. Circ. Sig. Process.* 69, 17–27 (2011)
9. Cheng, K.Y., Kumar, A.: Contactless Finger Knuckle Identification using Smartphones. In: Proc. of International Conference of the Biometrics Special Interest Group, BIOSIG (2012)
10. Zhang, L., Zhang, L., Zhang, D.: Finger Knuckle Print: A New Biometric Identifier. In: Proc. of ICIP 2009, pp. 1981–1984. IEEE (2009)
11. Zhang, L., Zhang, L., Zhang, D., Zhu, H.: Online finger-knuckle-print verification for personal authentication. *Pattern Recognition* 43, 2560–2571 (2010)
12. Zhang, L., Zhang, L., Zhang, D., Zhu, H.: Ensemble of local and global information for finger-knuckle print-recognition. *Pattern Recognition* 44, 1990–1998 (2011)
13. Kumar, A., Ravikanth, C.: Personal authentication using finger knuckle surface. *IEEE Trans. Information Forensics and Security* 4(1), 98–110 (2009)
14. Choraś, M., Kozik, R.: Knuckle Biometrics Based on Texture Features. In: Proc. of International Workshop on Emerging Techniques and Challenges for Hand-based Biometrics (ETCHB 2010). IEEE CS Press, Stambul (2010)
15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis* (2004)
16. Wang, J., Zha, H., Cipolla, R.: Combining interest points and edges for content-based image retrieval. In: Proceedings of the IEEE International Conference on Image Processing (2005)
17. Morales, A., Travieso, C.M., Ferrer, M.A., Alonso, J.B.: Improved finger-knuckle-print authentication based on orientation enhancement. *Electronics Letters* 47(6) (2011)

18. Hemery, B., Giot, R., Rosenberger, C.: Sift Based Recognition of Finger Knuckle Print. In: Proc. of Norwegian Information Security Conference, pp. 45–56 (2010)
19. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
20. Badrinath, G.S., Nigam, A., Gupta, P.: An Efficient Finger-Knuckle-Print Based Recognition System Fusing SIFT and SURF Matching Scores. In: Qing, S., Susilo, W., Wang, G., Liu, D. (eds.) ICICS 2011. LNCS, vol. 7043, pp. 374–387. Springer, Heidelberg (2011)
21. Goh, K.O.M., Tee, C., Teoh, B.J.A.: An innovative contactless palm print and knuckle print recognition system. *Pattern Recognition Letters* 31, 1708–1719 (2010)
22. Goh, K.O.M., Tee, C., Teoh, B.J.A.: Bi-modal palm print and knuckle print recognition system. *Journal of IT in Asia* 3 (2010)
23. Kumar, A., Zhou, Y.: Human Identification using Knuckle Codes. In: Proc. BTAS (2009)
24. Zhang, L., Li, H.: Encoding local image patterns using Riesz transforms: With applications to palmprint and finger-knuckle-print recognition. *Image and Vision Computing* 30, 1043–1051 (2012)
25. Zhang, L., Zhang, L., Zhang, D.: Finger-Knuckle-Print Verification Based on Band-Limited Phase-Only Correlation. In: Jiang, X., Petkov, N. (eds.) CAIP 2009. LNCS, vol. 5702, pp. 141–148. Springer, Heidelberg (2009)
26. Aoyma, S., Ito, K., Aoki, T.: Finger-Knuckle-Print Recognition Using BLPOC-Based Local block Matching, pp. 525–529. IEEE (2011)
27. Meraoumia, A., Chitroub, S., Bouridane, A.: Fusion of Finger-Knuckle-Print and Palmprint for an Efficient Multi-biometric System of Person Recognition. In: Proc. of IEEE ICC (2011)
28. Kumar, A., Prathyusha, K.V.: Personal Authentication Using Hand Vein Triangulation and Knuckle Shape. *IEEE Transactions on Image Processing* 18(9), 2127–2136 (2009)
29. Choraś, M., Kozik, R.: Contactless palmprint and knuckle biometrics for mobile devices. *Pattern Analysis and Applications* 15(1), 73–85 (2012)

Three-Stage Method of Text Region Extraction from Diagram Raster Images

Jerzy Sas and Andrzej Zolnierek

Abstract. In the paper the combined approach to the problem of text region recognition problem is presented. We focused our attention on the chosen case of text extraction problem from specific type of images where text is imposed over graphical layer of vector images (charts, diagrams, etc.). For such images we proposed three-stage method using OCR tools as some kind of feed-back in process of text region searching. Some experimental results and examples of practical applications of recognition method are also briefly described.

1 Introduction

Although *optical character recognition* (OCR) is a mature and widely used technique as far as clear text image is being processed, its performance is still low in case of text printed over textured background or the text is an element of the image consisting of many graphical elements not being text characters. In such cases, before the image is passed to OCR engine, the text must be extracted from background or the graphical elements not constituting the text must be removed. In this work we consider the specific case of text extraction form specific type of images where text is imposed over graphical layer of vector images. Typical examples of category of images are: plots, UML diagrams, organization charts, flow charts, electric schematic diagrams etc. Many works have been described in literature aimed on

Jerzy Sas
Wroclaw University of Technology, Institute of Applied Informatics,
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
e-mail: jerzy.sas@pwr.wroc.pl

Andrzej Zolnierek
Wroclaw University of Technology, Faculty of Electronics, Department of Systems and Computer Networks, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
e-mail: andrzej.zolnierek@pwr.wroc.pl

text extraction from real world pictures ([3]), where text appears over the complex background. In comparison with such kind of problems, text extraction from diagram-like images seems to be relatively easy. The experiments made with diagram images showed however that methods elaborated for text extraction from real word images, exhibit tendency to detect big number of "false positive" areas, i.e. the image fragments not being actually text areas are detected. The method presented in this article is an extension of approaches known from the literature, where multi-pass approach is used ([3], [4]). The first pass selects candidate fragments, possibly containing text, using edge-density based methods borrowed from the literature ([8]). Typically, it selects much more fragments than is finally expected. Then two subsequent passes refine this selection by applying more restrictive criteria based on geometric features of the fragment contents, specific for text images. The last selection stage uses results of OCR recognition ([2]). The relation of recognized text to the area aspect ratio as well as the analysis of the recognized string is used in order to further restrict the set of text area candidates. On the other hand in ([7]) connected-component-based character locating method is presented. Then in this method we can find several stages including color layers analysis, an aligning-and-merging-analysis and OCR applications. Another attempt is presented in ([1]) where mathematical morphology is used in text extraction before standard OCR tool is applied. In all above mentioned papers during the process of text region finding, before OCR module is applied some heuristic methods are proposed. Almost the same methodology is proposed here, but the novelty of this paper consists in field of application (charts, diagrams etc.) where appropriate algorithms should be applied. Moreover we propose using OCR module as some kind of feed-back in process of text region extraction.

The work described here is a part of wider project aimed on knowledge acquisition from electronic sources. Text extraction from images is used there a)for support of automatic images annotation (i.e. labeling images with keywords), b)finding documents which contents may help in answering user queries, c)extracting knowledge from images. Separation of text from remaining graphical elements of the image also makes easier the process of recognizing elements appearing on the image and finding determining relations between them. In order to achieve the last aim, it is necessary to precisely separate geometrical elements of the image (in the case of diagrams being considered here - line segments obtained as the result of the image vectorization) for text in the situation where graphics and text touch each other. This topic is also mentioned in this article.

The organization of the paper is as follows: after introduction the problem statement is presented in second section. Next, the description of three-stage method of text region extraction is presented. Every stage of our complex method is presented in subsequently subsections containing as well very well-known algorithms as well new ideas of this work. In the section 3 the results of empirical investigation are presented and in the end we conclude this paper.

2 Extraction Method of Region with Text

As stated in the introduction, the text extraction form the image is carried out for the sake of analysis of documents containing images. The complete electronic document (either scanned and stored as complete image, or structured, e.g. described in HTML) is first subdivided into structural elements (chapters, chapter titles, sections, paragraphs). The document analysis extracts images embedded in the document and stores them as raster images. These images are input to the text extraction procedure described here. Let us consider a digital raster image automatically extracted from the complete documents image containing some geometrical shapes (line segments, polygons, ellipses, vectorized symbols specific to diagram type etc.) and possibly containing areas where text is located. Text areas can be included inside the geometrical shapes, can be located out of them or can intersect edges of geometrical shapes. Our aim is to split the image into two layers: graphical and textual. The graphical layer contains all geometrical shapes not being elements of text. The textual one contains only elements belonging to texts included in the image. For the sake of OCR processing and in order to bind text to graphical shapes it is necessary to gather characters in the textual layer into *text regions*. The text region is the rectangular area tightly enclosing characters constituting the text string.

In the previous work the method of text regions extraction using texture features (local stochastic properties of visual features in pixels of the image) based on the algorithm described in ([8]) was applied. However, this method was excessively finding the region, which did not contain any text. In this work the methods based on texture properties are also used but they are augmented with additional features calculated for candidate regions which are rectangle covering the connected component. These features include the texture properties (in the form of a map like in ([4])) and the shape characteristics.

For the sake of further application in document analysis, the text extraction procedure creates the following artifacts:

- two binary images containing separated graphical and textual layer,
- the set of images corresponding to individual extracted text regions in two versions: as a binary image with white background and black foreground and as a simple fragment of the gray scale input image (including background),
- XML file describing the extracted text regions (position in the image, size, recognized text, link to the text region image).

The method proposed in this paper consists of three stages:

- first the set of *candidate regions* (which usually apart from the correct regions contains also the regions without the text) is determined and for such regions the feature vectors for pattern recognition task are calculated,
- in the second stage the candidate regions are classified as text regions or regions without the texts using the classifier (in which the above mentioned feature vector is used) based on simple decision tree, the regions recognized as not containing text are rejected from further considerations,

- on the last stage the elements classified as text region are recognized using standard OCR tool and the final decision about the region of interest is made using the likelihood of OCR decisions.

2.1 Determining of Candidate Regions

In our method it is assumed that texts appearing on the image are horizontally oriented. In order to extract vertically oriented texts proposed method is first applied to the original image. Then identified text regions are replaced by background color and the modified image is rotated by 90 degrees. Next the text extraction procedure is applied again to the rotated image. The ultimate text region set contains regions detected in the first or second pass.

The stage of candidate region elicitation consists of the following operations:

- conversion of the image into gray scale,
- filtering with the median filter to remove some noise being usually the result of scanning or lossy compression applied to the image,
- increasing the font sharpness by applying sharpening filter with Laplace mask,
- binarization of the image using Otsu method,
- removal of long line segments,
- finding connected components,
- removal of the connected components which are of one pixel size (in this way we can eliminate the remainder of the noise),
- the division of connected components into sets of big and small components, where the threshold value is experimentally chosen,
- removal of big connected components,
- merging the connected components into candidate regions basing on their geometric neighborhood.

The removal of big connected components is motivated by the assumption that the size of text in the image is limited and big graphical elements probably are not text characters. The size threshold for separation of big and small components is determined taking the image resolution x_{res}, y_{res} and the maximal assumed text size s_t (assumed to be 70 pixels).

$$f_s = \max(0.3 * \min(x_{res}, y_{res}), s_t) \quad (1)$$

Application of connected components as elements of candidate regions fails if the text is intersected by elements of other geometrical shapes. In order to avoid such cases as much as possible, we applied the additional step consisting in removal of long line segments. The lines removal procedure must not destroy the overlapping text image, because it would decrease the text detection accuracy as well as the accuracy of the OCR procedure. In order to achieve it, the binary raster image is skeletonized and vectorized, i.e., its representation as the set of line segments is created. Using vector representation of the image, basic 2D shapes like triangles,

ellipses and rectangles are recognized. The shape recognition procedure applies the set of geometric rules specific to each shape. All line segments belonging to recognized shapes are marked as excluded and assigned to the set E . Additionally, line segments not assigned to recognized shapes but which length exceeds the threshold defined by equation 1 are also assigned to E . Next, width of the lines in the original image corresponding to segments in E is approximated. Let $\pi(e)$ denotes the set of pixels of the skeleton line approximated by the line segment $e \in E$. The line width is evaluated as the doubled average distance between the skeleton line pixels $\pi(e)$ corresponding to the approximating line segment e to the closest background pixel:

$$w_e = 2 \frac{\sum_{p \in \pi(e)} \min_{b \in B} d(p, b)}{|\pi(e)|}, \quad (2)$$

where $d(p, b)$ denotes the Euclidean distance between pixels p and b and B is the set of background pixels. The idea of the line width approximation is shown in Fig. 1. Having line widths approximated, the lines in the set E are redrawn with the background color using their approximated widths. In this way, the long line segments disappear from the image. Unfortunately, also segments of text regions may be overwritten with the background color. To repair it, the foreground pixels located on the opposite sides of the line consisting of pixels $\pi(e)$ are connected by 1-pixel wide line segments drawn in the foreground color if their distance in less than $\sqrt{2}w_e$.

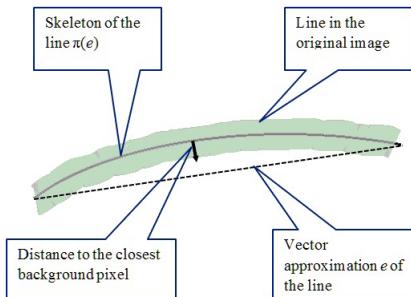


Fig. 1 Line width approximation by finding the closest background pixel

2.2 Classification of Candidate Regions

The stage of selection of candidate region consists of two steps:

- computing features for candidate regions,
- classification of candidate regions in order to eliminate the regions, for which the probability that they include the text is very small.

The regions elicited during the first stage of the whole process are separated only on the base of size and neighborhood of connected components in the binary image. In most cases, these regions in fact include the text fields but also very often the pseudo text regions which include other graphic elements of the image are created. That is why we need the selection process in which we can reject the region with small probability of text inclusion basing on another more complicated criteria. We applied typical approach using two-class pattern recognition, where the classes are: "the area includes text" and "the area does not include text". For candidate regions we determined the following features:

- foreground pixels coverage factor $f_{FB} = n_{FG}^{(a)} / (x_{res}^{(a)} * y_{res}^{(a)})$, where $n_{FG}^{(a)}$ is the number of foreground pixels in the region a and $x_{res}^{(a)}, y_{res}^{(a)}$ define the horizontal and vertical resolution of the region,
- the ratio of average line thickness to the height of the corresponding region (it allows for rejection of such regions in which this ratio is small what happens typically for graphical elements),
- the difference between the average background brightness and the average foreground pixels brightness (it allows for rejection of regions which are rather textures areas),
- density of connected components, i.e. the ratio of the connected components number to the area width / height aspect ratio

$$f_{CD} = \frac{n_C^{(a)}}{x_{res}^{(a)} / y_{res}^{(a)}} \quad (3)$$

where $n_C^{(a)}$ is the number of connected components in the area a ,

- the average width of consistency component calculated with respect to the height of the field (we assumed that consistency components correspond to the individual graphic characters, so their width should be less or equal to the height of the field),
- the average distance of foreground pixels to the regression line $y = ax + b$ (obtained by the Least Square Method) for the set of foreground pixels with respect to the height h of the field (it allows for rejecting the regions containing the simple shapes like a line segments):

$$f_{AD} = \frac{\sum_{i=1}^n |y_i - ax_i - b|}{nh} \quad (4)$$

- maximal distance of foreground pixels to the regression line determined for the set foreground pixels, divided by the height of the region,
- the average value from gradient map obtained as a sum of directional detectors of edges obtained by the method described in ([4]).

The classification rules for text regions or for the regions without the text are of the form of comparison the feature values with their thresholds. Next these predicates are used in a classifier based on simply decision tree. In such manner, a part of

candidate regions is rejected and the rest of them is subsequently processed using OCR tool.

2.3 Verification of Text Recognition with OCR Tool

The regions selected during the second stage of classification process now are analyzed by OCR tool. In our work, taking into account its accuracy proved by other researchers, the Tesseract OCR tool was chosen. Unfortunately, Tesseract as well as many other robust OCR systems does not provide the evaluation of confidence of the recognition results. Because this knowledge is important in our system, then we try to estimate the confidence taking into account the contents of the recognized text string. One of the obvious approaches may use Levenshtein distance ([6]) of the recognized string to the closest sequence of words from the fixed dictionary. This approach hardly can be applied in the problem being considered here because it is almost impossible to build the dictionary of symbols appearing in a diagram. We however observed that in the case of the falsely selected regions, OCR applied to it produces the sequence of characters mainly containing symbols not being letters nor digits (*dirty symbols*). The set of dirty symbols consists of semi-graphic characters, punctuation marks " .,:;~! " and operators " #, -, +, /, *, &, ^" (if they appear side by side). If the sequence recognized by OCR contains plenty of such symbols (especially out of the set of typical punctuation marks like "...;~" then the result is not reliable, and most probably the region passed to OCR is not a true text region. As a measure of the likelihood for the recognized inscription, the following formula is adopted. The leading and trailing white characters: spaces and tabs, are not taken into account:

$$\eta = \max(0, \frac{l - n_d - n_p - 0.2n_L}{l}) \quad (5)$$

where: l is the length of the recognized string, n_d is the number of dirty characters contained in the string not being typical punctuation marks nor operators, n_p is the number of punctuation marks and operators not being neighbors of other dirty characters and n_L is the sum of minimal Levenshtein distances computed for subsequences of the recognized string consisting exclusively of letters or digits. By minimal Levenshtein distance with respect to the dictionary, we mean the distance to the closest word in the dictionary. In order to calculate n_L the complete dictionary is necessary for language which is used in the inscriptions on the image. In some types of diagram (for example in UML diagrams) inscriptions containing the identifiers which are not the words from the vocabulary are very likely to appear. That is why this factor has small significance in the formula (5). The small value of likelihood coefficient denotes either that the quality of image is insufficient for correct recognition or the region transferred to the OCR module does not contain pure text. In both cases it seems to be a reasonable cause for rejecting the field as real text region.

Additionally for evaluation of text likelihood one can use the comparison between expected and real width of text field. We noticed that in the case of false classification of the field as text region, there are significant differences between the

real width of region under consideration and its expected width calculated as a sum of average widths of its component characters. These width are computed with respect to the high of the text field. By taking into account above mentioned relation, the recognition accuracy evaluation can be calculated as:

$$\lambda = \sqrt{2^{-|log_2(w/h_{est})|}} \quad (6)$$

where w_{est} is the estimated width of text recognized by OCR module, obtained by summing average ratio coefficients (width/height) for all characters appearing in recognized inscription. Finally the likelihood of recognition is determined as product $\eta\lambda$. During experiments we found that the region should be rejected if calculated likelihood is less than 0.3 .

3 Experimental Results

In tests we used seven typical class of diagrams: organization charts, UML sequence diagrams, pie charts, line charts, UML use case diagrams, flow charts, column charts. As a pattern the text field was chosen. We defined the text field as a geometrical closed region containing literary marks creating at least either one word of length not less than three or a number consisting of at least one digit. We did not expect recognition of isolated very short words of length less than three letters because they do not contain any useful information in document searching process and moreover it is very difficult to discriminate them from another small graphic elements of the image. As a measure as usual accuracy and recall were chosen. In the description of empirical studies the following notation was introduced:

- n_{tp} - the number of text fields, which were correctly found on the image,
- n_{tn} - the number of text fields, which were on the image but we did not find them correctly,
- n_{fp} - the number of subimage, which were recognized as a text fields but in fact they are not (we did not count the region including only one letter or pair of letters, because in reality we deal with text data).

In consequence the total number of words on the images is equal to $n_{tp} + n_{tn}$ while the total number of regions indicated by our method is equal to $n_{tp} + n_{fp}$. The accuracy and the recall were determined as follows:

$$f_K = \frac{n_{tp}}{n_{tp} + n_{tn}}, \quad f_P = \frac{n_{tp}}{n_{tp} + n_{fp}} \quad (7)$$

The result of empirical investigation concerning the accuracy and the recall of text extraction are presented in the table 1.

Let us notice that our investigation was focussed on accuracy of text region finding but not on the accuracy of OCR.

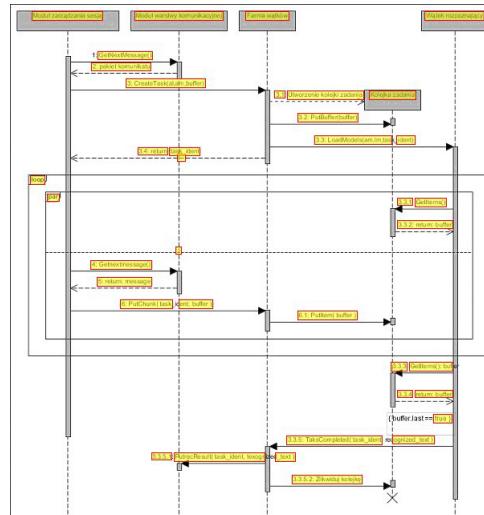
Table 1 Accuracy and recall of text extraction for different kind of diagrams

Category	n_{tp}	n_{tn}	n_{fp}	f_K	f_P
Organization charts	1595	175	77	0.9011	0.9539
UML sequence diagrams	291	37	36	0.8872	0.8899
Pie charts	143	14	3	0.9108	0.9795
Line charts	1177	185	67	0.8642	0.9461
UML use case diagrams	112	3	8	0.9739	0.9333
Flow charts	442	40	3	0.9170	0.9933
Column charts	703	72	26	0.9071	0.9643
Total	4463	526	220	0.8946	0.9530

The accuracy of our method is the greatest in the case of flow diagrams and block schemes and the least for schemes of UML sequences. The accuracy is decreasing in these case when:

- there exist parts of text which intersect with the graphic elements of the image,
- there exist small graphic elements of shapes similar to literary marks,
- there exist text elements written with the small font (the accuracy rapidly is decreasing when the font is of height less than seven pixels).

We can also notice that the long inscriptions, which are the most important in text recognition on the image, were almost always correctly found. In the case of inscription written with the small letters the OCR module plays the crucial role as well in text region finding as in final text recognition. In this work external OCR module

**Fig. 2** The results of text extraction on the UML sequence diagram

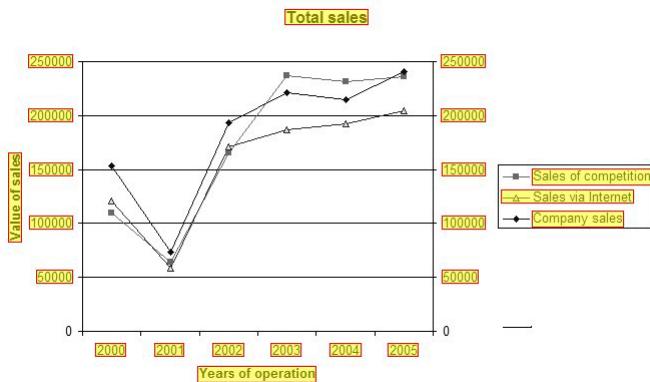


Fig. 3 The results of text extraction on the line chart

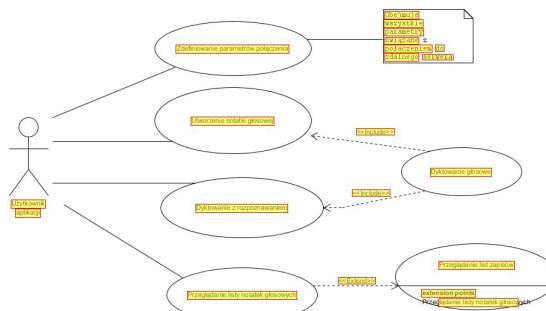


Fig. 4 The results of text extraction on the UML use case diagram

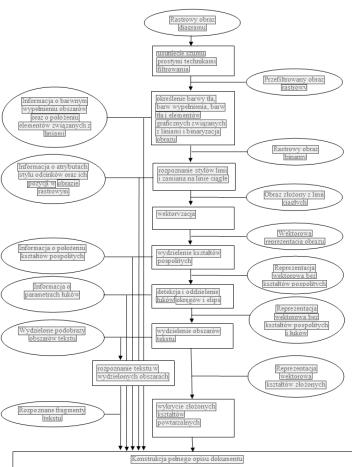


Fig. 5 The results of text extraction on the flow chart

was applied and its accuracy had the greatest influence for accuracy of whole method in the case mentioned above.

The results of text region extraction on the images of chosen investigated category were presented on the pictures 2 to 6. Additionally on the picture 7 the result of application of proposed method for drawings which is not a chart or a diagram is presented. The fields identified as the text regions are enclosed by rectangular boxes filled with gray shade.

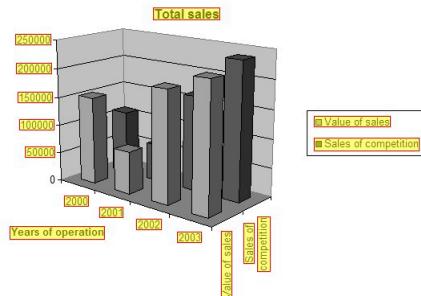


Fig. 6 The results of text extraction on the column chart

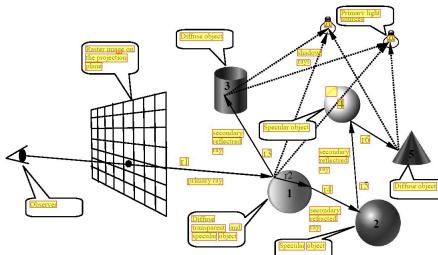


Fig. 7 The results of text extraction on the image which is not a diagram

4 Conclusions

In this paper we have focused our attention on the three-stage algorithm of text extraction from the images in case of text printed over textured background or the text is an element of the image consisting of many graphical elements not being text characters. In this work we consider the specific case of text extraction from specific type of images where text is imposed over graphical layer of vector images. Typical examples of category of images are: plots, UML diagrams, organization charts, flow charts, electric schematic diagrams etc. For such images we adopted several heuristic algorithms allowing finding the text region. We also applied some

kind of feed-back in this process using OCR tool taking into account its estimated accuracy.

Presented experimental results concerning proposed method imply the following conclusions:

- this algorithm seems to be an interesting alternative to another algorithms for text extraction,
- the accuracy and the recall of our method is acceptable for practical application,
- it can be effectively used also in the case when do not deal with diagrams,

although more practical investigations should be made. In particular we plan to apply our method to ICDAR 2003 benchmark base although our method is diagram-oriented. Moreover, it seems that applying own OCR procedure or procedure in which the support vectors for decision making are known could improve the quality of whole system because these factors play important role there.

Acknowledgements. This work has been in part supported by the Innovative Economy Programme project POIG.01.01.02-14-013/09.

References

1. Babu, G., Srimaiyee, P., Srikrishna, A.: Text extraction from heterogenous images using mathematical morphology. *Journal of Theoretical and Applied Information Technology*, 39–47 (2010)
2. Epshtain, B., Ofek, E., Wexler, Y.: Detecting text in natural scenes with stroke width transform. In: Proc of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 2963–2970 (2010)
3. Gllavata, J., Ewerth, R., Freisleben, B.: Proc. of 3rd International Symposium on Image and Signal Processing and Analysis, ISPA 2003, Rome, pp. 611–616 (2003)
4. Liu, X., Samarabandu, J.: Multiscale edge-based text extraction from complex images. In: Proc. of IEEE International Conference on Multimedia and Expo., pp. 1721–1724. IEEE Computer Society (2006)
5. Marti, U.V., Bunke, H.: Using a Statistical Language Model to Improve the Performance of an HMM-Based Cursive Handwriting Recognition System. *Int. Journ. of Pattern Recognition and Artificial Intelligence* 15, 65–90 (2001)
6. Paleo, B.: Levenshtein distance: Two Application in Data Base Record Linkage and Natural Language processing. LAP Lambert Academic Publishing (2010)
7. Wang, K., Kangasb, J.: Character Location in Scene Images from Digital Camera. *Pattern Recognition* 36, 2287–2299 (2003)
8. Wu, V., Manamatha, R., Riseman, E.: Finding text in images. In: Proc. of the Second ACM International Conference on Digital Libraries, pp. 3–12 (1997)

Part V

Medical Applications

Time Series of Fuzzy Sets in Classification of Electrocardiographic Signals

Jacek M. Leski and Norbert Henzel

Abstract. A way of an application of time series of fuzzy sets to classification of QRS complexes of ECG signal for selected averaging of this signal is presented. After the formulation of the problem the notion of time series of fuzzy sets is recalled. The time series of fuzzy sets are created on the basis of the original noisy signal. The parameters of successive fuzzy sets are used as a feature vector for a classifier. In the presented paper, the ℓ_2 -regularized iteratively reweighted least squares classifier and its kernel version are used. The MIT-BIH annotated ECG database is used in the experiments. The multi-fold cross-validation procedure using 100 pairs of learning and testing subsets are applied to validate the classification results. The obtained results (generalization error less than 1%) are very promising.

Keywords: Biomedical Signals Classification, Fuzzy Sets, Time Series.

1 Introduction

Fuzzy sets have been created to describe and to process an imprecise information, that can easily be expressed in a natural language prepositions, that are friendly to humans/experts. The biomedical signals are difficult to describe using traditional methods of mathematical modeling. Since 1996, one can observe attempts to use concepts of fuzzy sets in the description and the analysis of biomedical signals.

Jacek M. Leski

Institute of Medical Technology and Equipment, Poland Institute of Electronics,
Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
e-mail: jacekl@itam.zabrze.pl, jleski@polsl.pl

Norbert Henzel

Institute of Electronics, Silesian University of Technology, Akademicka 16,
44-100 Gliwice, Poland
e-mail: rhenzel@polsl.pl

In [1] the concept of the so-called fuzzy signal has been proposed and has been used to entropy and energy measures of fuzziness in ECG signals processing. The idea of representation of ECG signals using a sequence of information granules was introduced in [2]. Paper [4] presents unification of the methods of description ECG signals using fuzzy sets, called the time series of fuzzy sets. The ECG signal is processed in a moving window in time domain and transformed into a time series of membership functions, which describe fuzzy sets. In special cases the method of fuzzy signal and sequence of information granules can be considered.

The averaging method of the ECG signal in time domain requires classified QRS complexes, because the use of only one complex of different morphology for averaging results in serious errors of the method. The classification of QRS complexes as one of the two classes: the normal (dominant) class and the ventricular one, causes some problems for noisy signals.

The main goal of this work is to show that the concept of the time series of fuzzy sets can be used for classification of noisy QRS complexes. The next goal is to investigate the generalization ability of the QRS ℓ_2 -regularized iteratively reweighted least squares classifier and its kernel version [6] using the real-world benchmark MIT-BIH [7] database.

2 Time Series of Fuzzy Sets

In [4] it was noted that the method based on fuzzy signal and method using information granules have similarities, of which the most important are:

- both methods convert the sequence of signal samples contained in a time window into the set of fuzzy triangular or trapezoidal membership functions;
- time sequence of the created fuzzy sets has a linguistic interpretation and can be used for classification of signals.

However, these methods also differ considerably:

- subsequent time windows overlap for the method based on fuzzy signal, whereas for information granules subsequent time windows are adjacent;
- the membership function for information granules are created by the solution of the optimization problem, whereas for the fuzzy signal a fuzzy set is created by ordering of the signal samples;
- the sequence of resulting fuzzy sets are used in two different ways: to determine entropy or energy measures of fuzziness, or to create a fuzzy vocabulary.

An idea to unify both of the above-mentioned methods uses the concept of fuzzy time series. Let us consider the n th signal sample and the window formed from $2k+1$ samples of the original signal:

$$\{g(n-k), g(n-k+1), \dots, g(n), \dots, g(n+k-1), g(n+k)\}, \quad (1)$$

were $g(n)$ denotes discrete signal. After the ordering operation with respect to the amplitude of the samples, we obtain:

$$g_{(1)}(n) \leq g_{(2)}(n) \leq \cdots \leq g_{(2k+1)}(n) \quad (2)$$

The above operations are performed for every j th discrete time index windows location, i.e. $\cdots, g(n-2j), g(n-j), g(n), g(n+j), g(n+2j), \dots$. For $j = 1$ the fuzzy signal method is obtained, while for $j = 2k$, the method based on information granules is derived.

There are many ways to create a membership function of a fuzzy set on the basis of the ordered signal samples. However, they are characterized by high computational complexity. In [4] it was proposed a method to create a membership function with low computational effort, while not susceptible to noises associated with ECG signal, i.e.: muscle noise, baseline wander and powerline interference.

It is proposed to take gaussian membership functions with the values of the parameters calculated on the basis of the samples contained in a given window (1):

$$\mu_{A(n,k)}(g) = \exp \left\{ -\frac{[g - c(n)]^2}{s(n)} \right\}, \quad (3)$$

where $A(n,k)$ stands for fuzzy set for discrete time n and window length $(2k + 1)$; $c(n)$ and $s(n)$ denotes central value and dispersion, respectively. The central value for n th discrete time is calculated as the median value of signal samples in the analysis window: $c(n) = g_{(k+1)}(n)$ (see (2)). Denote the deviation of samples from the central value as $r_i(n) = g_{(i)}(n) - c(n)$, where $i = 1, 2, \dots, 2k + 1$, and using robust estimator of dispersion we obtain [4]:

$$s(n) = 1.4826 \left(1 + \frac{5}{2k} \right) \sqrt{\text{med}_i r_i(n)^2}, \quad (4)$$

where 'med' stands for the median operation. In practical application Gaussian membership function can be approximated using a simpler, triangular membership function. The triangular membership function is represented as a symmetrical triangle with base length equal to $2p(n)$, where, based on [5], $p(n) = 2.375s(n)$, and a maximum placed at $c(n)$. For each QRS complex the time series of fuzzy sets is created using the following parameter values $k = 10$ and $j = 22$. Figures 1 and 2 present examples of the time series of fuzzy sets for normal and ventricular heartbeats, respectively. Feature vector for the i th QRS complex takes the form:

$$\mathbf{x}_i^\top \triangleq [c(FP_i - 6j), \dots, c(FP_i), \dots, c(FP_i + 7j), \\ s(FP_i - 6j), \dots, s(FP_i), \dots, s(FP_i + 7j)], \quad (5)$$

where FP_i denotes the fiducial point for the i th QRS complex.

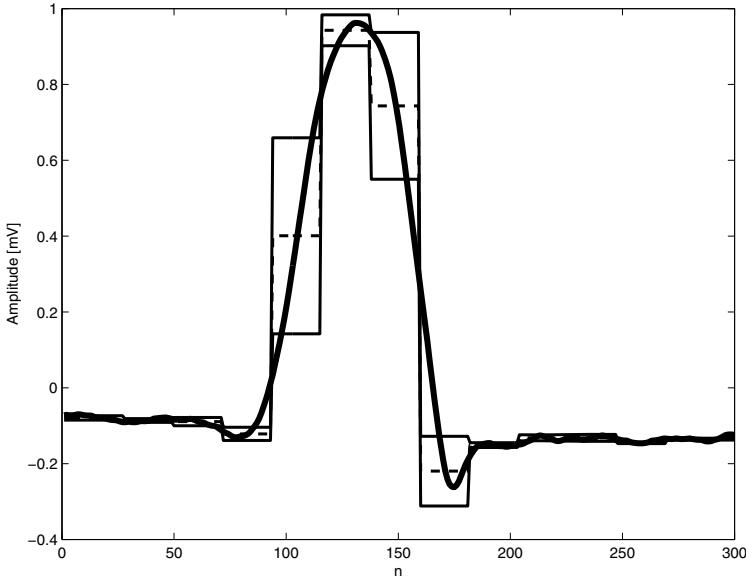


Fig. 1 An example of QRS complex taken from the normal class (bold line) and corresponding time series of triangular fuzzy sets (dotted line for center value; normal line for triangle base)

3 Classification Methodology – Linear Case

The classifier is designed on the basis of a training set, $\mathcal{T}_{\mathcal{R}}^{(N)} = \{(\mathbf{x}_1, \theta_1), (\mathbf{x}_2, \theta_2), \dots, (\mathbf{x}_N, \theta_N)\}$, where N is QRS complexes cardinality, and each independent feature vector, containing the parameters of fuzzy sets $\mathbf{x}_i \in \mathbb{R}^t$ has a corresponding dependent datum $\theta_i \in \{+1, -1\}$, which indicates its assignment to one of two classes, ω_1 (normal) or ω_2 (ventricular): $\theta_i = +1$ iff $\mathbf{x}_i \in \omega_1$ and $\theta_i = -1$ iff $\mathbf{x}_i \in \omega_2$. Defining the augmented pattern vector $\mathbf{x}'_i = [\mathbf{x}_i^\top, 1]^\top$, we seek a weight vector $\mathbf{w} = [\tilde{\mathbf{w}}^\top, w_0]^\top \in \mathbb{R}^{t+1}$, such that

$$d(\mathbf{x}_i) \triangleq \mathbf{w}^\top \mathbf{x}'_i = \tilde{\mathbf{w}}^\top \mathbf{x}_i + w_0 \begin{cases} \geq 0, & \mathbf{x}_i \in \omega_1, \\ < 0, & \mathbf{x}_i \in \omega_2, \end{cases} \quad (6)$$

where $d(\mathbf{x}_i)$ is a linear discrimination function.

If we multiply by -1 all patterns of the training set that are members of ω_2 class, then (6) can be rewritten in the form $\theta_i \mathbf{w}^\top \mathbf{x}'_i \geq 0$, for $i = 1, 2, \dots, N$. According to the statistical learning theory [9] the safer approach is to seek the vector \mathbf{w} , such that

$$\theta_i \mathbf{w}^\top \mathbf{x}'_i \geq 1 \quad (7)$$

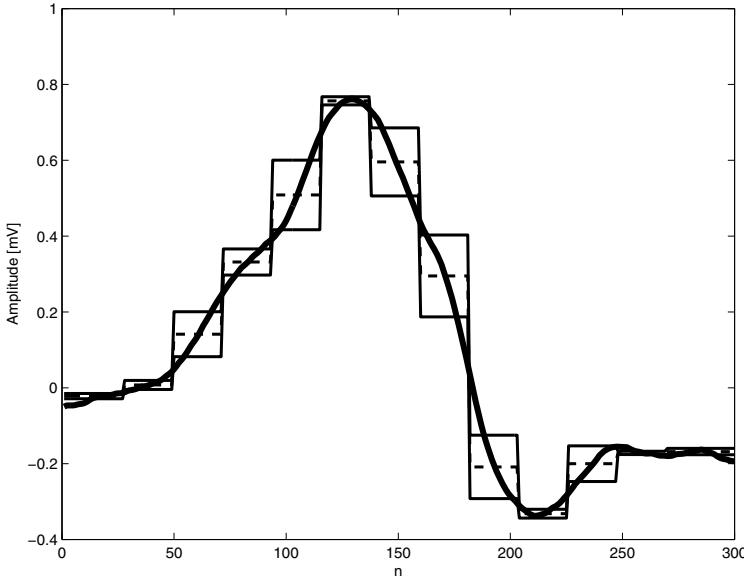


Fig. 2 An example of QRS complex taken from the ventricular class (bold line) and corresponding time series of triangular fuzzy sets (dotted line for center value; normal line for triangle base)

and minimizing $\tilde{\mathbf{w}}^\top \tilde{\mathbf{w}}$. Let \mathbf{X} be the $N \times (t+1)$ matrix

$$\mathbf{X}^\top \triangleq [\theta_1 \mathbf{x}'_1, \theta_2 \mathbf{x}'_2, \dots, \theta_N \mathbf{x}'_N]. \quad (8)$$

Then inequalities (7) can be rewritten in the matrix form $\mathbf{X}\mathbf{w} \succeq \mathbf{1}$, where $\mathbf{1}$ denotes the vector with all entries equal to 1 and the symbol \succeq stands for componentwise inequality. The error vector is defined as $\mathbf{e} = \mathbf{X}\mathbf{w} - \mathbf{1}$. We seek vector \mathbf{w} by the following minimization [6]

$$\min_{\mathbf{w} \in \mathbb{R}^{t+1}} J(\mathbf{w}) \triangleq \sum_{i=1}^N \frac{h_i}{2} \mathcal{L}(\theta_i \mathbf{w}^\top \mathbf{x}'_i - 1) + \frac{\tau}{2} \tilde{\mathbf{w}}^\top \tilde{\mathbf{w}}, \quad (9)$$

where $\mathcal{L}(\cdot)$ stands for a loss function used to approximation of misclassification error, h_i is a weight corresponding to the i th pattern. The second term is related to the maximization of the margin of separation (classifier complexity minimization). Parameter $\tau > 0$ controls the trade-off between the complexity of classifier and the amount up to which errors are tolerated.

If we choose the quadratic loss function then in matrix notation (9) takes the form

$$\min_{\mathbf{w} \in \mathbb{R}^{t+1}} J(\mathbf{w}) \triangleq \frac{1}{2} (\mathbf{X}\mathbf{w} - \mathbf{1})^\top \mathbf{H} (\mathbf{X}\mathbf{w} - \mathbf{1}) + \frac{\tau}{2} \tilde{\mathbf{w}}^\top \tilde{\mathbf{w}}, \quad (10)$$

where the matrix $\mathbf{H} = \text{diag}(h_1, h_2, \dots, h_N)$. Indeed, this loss function is symmetric – the patterns equally distant from the edge of the region of separation, independently they lie on the right or on the wrong side has the same penalty. To obtain asymmetric loss function, it is proposed in paper [6], that the weights h_i may be used. If we set $h_i = 0$ for $e_i \geq 0$ and 1 otherwise, then only patterns lie on the region of separation are penalized. However, to check the sign of e_i we must known vector \mathbf{w} . Thus, criterion function (10) should be minimized by iteratively reweighting scenario. Let us denote \mathbf{w} , \mathbf{H} and \mathbf{e} in k th iteration as $\mathbf{w}^{(k)}$, $\mathbf{H}^{(k)}$ and $\mathbf{e}^{(k)}$, respectively. Criterion function (10) for k th iteration takes the form [6]

$$\begin{aligned} \min_{\mathbf{w}^{(k)} \in \mathbb{R}^{t+1}} J^{(k)}(\mathbf{w}^{(k)}) &\triangleq \frac{1}{2} \left(\mathbf{X}\mathbf{w}^{(k)} - \mathbf{1} \right)^T \mathbf{H}^{(k)} \left(\mathbf{X}\mathbf{w}^{(k)} - \mathbf{1} \right) \\ &+ \frac{\tau}{2} \left(\tilde{\mathbf{w}}^{(k)} \right)^T \tilde{\mathbf{w}}^{(k)}, \end{aligned} \quad (11)$$

where

$$h_i^{(k)} = \begin{cases} 0, & e_i^{(k-1)} \geq 0, \\ 1, & e_i^{(k-1)} < 0, \end{cases} \quad (12)$$

$$\mathbf{e}^{(k-1)} = \mathbf{X}\mathbf{w}^{(k-1)} - \mathbf{1}. \quad (13)$$

Thus, to minimize the criterion function for the k th iteration the weights are obtained using (12), (13) and the result of optimization of the criterion function for the previous iteration. To start this sequential optimizations we set weights in the 0th iteration as $h_i = 1$ for all i . Above minimization is called as Iteratively Reweighted Least Square (IRLS) method with control the complexity of a solution [6]. The loss function obtained in (11), (12) is called Asymmetric SQuaRe (ASQR) function. Classifier design methods need to be robust. Thus, for a better approximation of misclassification error, the absolute error function may be easily obtained by taking [6]:

$$h_i^{(k)} = \begin{cases} 0, & e_i^{(k-1)} \geq 0, \\ \frac{1}{|e_i^{(k-1)}|}, & e_i^{(k-1)} < 0. \end{cases} \quad (14)$$

In this case, reweighting is used for both asymmetrization (relaxation) and changing the loss function. The loss function obtained in (11), (16) is called Asymmetric LINear (ALIN) function. Many other loss functions may be easily obtained [6]:

- Asymmetric HUBer (AHUB)

$$h_i^{(k)} = \begin{cases} 0, & e_i^{(k-1)} \geq 0, \\ 1, & -1 \leq e_i^{(k-1)} < 0, \\ \frac{1}{|e_i^{(k-1)}|}, & e_i^{(k-1)} < -1. \end{cases} \quad (15)$$

- SIGmoidal (SIG)

$$h_i^{(k)} = \frac{1}{\left|e_i^{(k-1)}\right|^2 \left\{1 + \exp\left[\alpha\left(e_i^{(k-1)} + 1\right)\right]\right\}}. \quad (16)$$

- Asymmetric SIGmoidal-Linear (ASIGL)

$$h_i^{(k)} = \frac{1}{\left|e_i^{(k-1)}\right| \left\{1 + \exp\left[\alpha\left(e_i^{(k-1)} + 1\right)\right]\right\}}. \quad (17)$$

- Asymmetric LOGarithmic (ALOG)

$$h_i^{(k)} = \begin{cases} 0, & e_i^{(k-1)} \geq 0, \\ \frac{\log\left(1 + |e_i^{(k-1)}|^2\right)}{|e_i^{(k-1)}|^2}, & e_i^{(k-1)} < 0. \end{cases} \quad (18)$$

- Asymmetric LOG-Linear (ALOGL)

$$h_i^{(k)} = \begin{cases} 0, & e_i^{(k-1)} \geq 0, \\ \frac{\log\left(1 + |e_i^{(k-1)}|^2\right)}{|e_i^{(k-1)}|}, & e_i^{(k-1)} < 0. \end{cases} \quad (19)$$

The condition for optimality of (11) for the k th iteration, as a system of linear equations, is obtained by its differentiating with respect to \mathbf{w} and setting the result equals to zero

$$\left(\mathbf{X}^\top \mathbf{H}^{(k)} \mathbf{X} + \tau \tilde{\mathbf{I}}\right) \mathbf{w}^{(k)} = \mathbf{X}^\top \mathbf{H}^{(k)} \mathbf{1}, \quad (20)$$

where $\tilde{\mathbf{I}}$ is the identity matrix with the last element on the main diagonal equals zero. In [6] it is proposed to solve (20) using conjugate gradient method.

The iteratively reweighted least square error minimization procedure for classifier design can be summarized in the following steps:

1. Fix $\tau > 0$ and $\mathbf{H}^{(0)} = \mathbf{I}$. Set the iteration index $k = 0$.
2. obtain $\mathbf{w}^{(k)}$ by using conjugate gradient method to (20).
3. $\mathbf{e}^{(k)} = \mathbf{X}\mathbf{w}^{(k)} - \mathbf{1}$.
4. $\mathbf{H}^{(k+1)} = \text{diag}\left(h_1^{(k+1)}, h_2^{(k+1)}, \dots, h_N^{(k+1)}\right)$, where $h_i^{(k+1)} = f\left(e_i^{(k)}\right)$, for $i = 1, 2, \dots, N$, and $f(\cdot)$ stands for selected loss function.
5. if $k > 1$ and $\|\mathbf{w}^{(k)} - \mathbf{w}^{(k-1)}\| < 10^{-3}$, then stop
else $k \leftarrow k + 1$, go to 2.

4 Classification Methodology – Nonlinear Case

Among nonlinear classifiers the so-called kernel-based methods are of special interest in last years [8]. The methods may be motivated by the Cover's theorem on separability of patterns. This theorem states that nonlinearly separated patterns in the input space can be linearly separable in a new space obtained by nonlinear mapping of the original one, if the dimensionality of new space is high enough [3]. Let $\Phi : \mathbf{x} \in \mathbb{R}^t \mapsto \Phi(\mathbf{x}) \in \mathcal{F}$ be a nonlinear transformation of the input vectors \mathbf{x} into a feature space \mathcal{F} , which may be high- or even infinite-dimensional. Taking into account that K may be decomposed $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^\top \Phi(\mathbf{x}_j)$ means that we design a linear classifier in the feature space \mathcal{F} , i.e. a nonlinear one in the original input space. The most commonly used kernel function is gaussian:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\chi \|\mathbf{x} - \mathbf{x}_i\|^2\right); \quad \chi \in \mathbb{R}_+, \quad (21)$$

where χ denotes parameter. Defining $(N \times N)$ -dimensional kernel matrix

$$\mathcal{K} = [\theta_i \theta_j K(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^N, \quad (22)$$

$\boldsymbol{\Gamma}^\top = [\gamma_0, \gamma_1, \dots, \gamma_N]^\top$, $\boldsymbol{\Theta}^\top = [\theta_1, \theta_2, \dots, \theta_N]^\top$ and $\widetilde{\mathcal{K}} = [\boldsymbol{\Theta} \mathcal{K}]$, minimized criterion for weighted (using $\mathbf{H}^{(k)}$) least square loss takes the form

$$\begin{aligned} \min_{\boldsymbol{\Gamma}^{(k)} \in \mathbb{R}^{N+1}} J^{(k)}(\boldsymbol{\Gamma}^{(k)}) &= \frac{\tau}{2} \left(\boldsymbol{\Gamma}^{(k)}\right)^\top \widetilde{\mathcal{I}}^\top \mathcal{K} \widetilde{\mathcal{I}} \boldsymbol{\Gamma}^{(k)} \\ &+ \frac{1}{2} \left(\widetilde{\mathcal{K}} \boldsymbol{\Gamma}^{(k)} - \mathbf{1}\right)^\top \mathbf{H}^{(k)} \left(\widetilde{\mathcal{K}} \boldsymbol{\Gamma}^{(k)} - \mathbf{1}\right), \end{aligned} \quad (23)$$

where matrix $\widetilde{\mathcal{I}} = [\mathbf{0} \ \mathbf{I}]$ is used to unregularize the bias term. The decision function of the classifier for input pattern \mathbf{x} can be represented as

$$d(\mathbf{x}) = \sum_{i=1}^N \theta_i \gamma_i K(\mathbf{x}, \mathbf{x}_i) + \gamma_0, \quad (24)$$

where $\{\gamma_i\}_{i=0}^N$ are parameters of the classifier obtained in the process of training. The training is done using an algorithm similar to the one presented for linear case. For details see [6].

5 Numerical Experiments and Discussion

All experiments were done on Hewlett-Packard HP Compaq dx7300 Intel Core 2 CPU 6300 @ 1.86 GHz with 1GB RAM, running Windows XP (Service Pack 2) and MATLAB 6.5 environment. For experiments the well-known benchmark database MIT-BIH [7] was used. This database includes 48 of 31 min. long ECG recordings

with annotations. From 82182 QRS complexes marked as Normal (N) and Ventricular (V), 105 Ns and 105 Vs was randomly selected for further analysis. This set of QRS complexes were interpolated to sampling rate equal to 2000 Hz (using MATLAB 'interp' function), and then, transformed into time series of fuzzy sets, and next, were randomly partitioned into 100 pairs of training (35 Ns and 35 Vs QRSSs) and testing (70 Ns and 70 Vs QRSSs) sets. The structure of experiments is as follows: for database 10% of splits were used to estimate the parameters (regularization parameter and kernel parameter). These parameters were used to 100-fold cross validation. The average and standard deviation of generalization error were used to compare the performance with various approximations of misclassification error in both linear and nonlinear classifiers.

6 Results

In experiments the regularization parameter τ was in the range from 0.1 to 2.5 (step 0.01). The Gaussian kernel (21) was used. The kernel parameter χ was selected from the set $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.5\}$. Table 1 shows the average generalization performance and standard deviation. Clearly, the classification results differs significantly for the linear and the nonlinear classifier. In general, the mean of classification error is about three-fold more significant for the linear classifier with respect to the nonlinear one. In the case of standard deviation, the difference is about two-fold in favor for the nonlinear method. In the case of the linear classifier, the least mean value of classification error is obtained for the ASQR function. Slightly higher mean value is obtained for the ALOGL function. In the case of standard deviation the situation is opposite. The smallest standard deviation is obtained for the ALIN function. For the nonlinear classifier the picture is quite different. The best classification quality is obtained for the ASQR, the AHUB and the ALOG functions.

Table 1 Comparison between various approximations of misclassification error for linear and nonlinear (kernel) IRLS classifier on the MIT-BIH dataset

Loss	Classification error					
	Linear		Nonlinear			
	Mean \pm St.Dev	τ	Mean \pm St.Dev	τ	χ	
ASQR	2.893 ± 1.970	1.27	0.679 ± 0.821	0.15	0.01	
ALIN	3.229 ± 1.635	1.91	1.043 ± 0.840	0.17	0.10	
ASIG	3.543 ± 1.805	1.95	1.043 ± 0.840	0.21	0.10	
ASIGL	3.529 ± 1.816	1.95	1.043 ± 0.840	0.21	0.10	
AHUB	3.250 ± 1.834	1.83	0.679 ± 0.821	0.12	0.01	
ALOG	3.164 ± 1.980	0.14	0.679 ± 0.821	0.11	0.01	
ALOGL	2.921 ± 1.858	2.41	1.614 ± 0.736	0.35	0.01	

Second best classification errors are get for the ALIN, the ASIG and ASIGL functions. The worst mean of classification errors is received for the ALOGL function; in contrast, this function permits to get the best standard deviation of all examined cases.

7 Conclusions

In the work, the ℓ_2 -regularized iteratively reweighted least squares classifier and its kernel version are used to classify noisy QRS complexes to one of the two classes, i.e. the normal and the ventricular. The time series of fuzzy sets are created on the basis of the original noisy signal. The parameters of successive fuzzy sets are used as a feature vector for a classifier. An extensive experimental analysis on MIT-BIH benchmark database showing that proposed method are good alternative, in terms of generalization performance, with respect to the methods known from the literature. For the linear classifier the best performance is observed on the ASQR and the ALOGL approximation of misclassification error. For the kernel version of classifier the best are the AQRS, the AHUB and the ALOG approximations. Additionally, feature vector used to classification, based of fuzzy sets parameters, may enable sharing the knowledge with human experts (physicians). The obtained classification quality seems to confirm, that the classification of the time series of fuzzy sets is appropriate for application in selective signal averaging.

References

1. Czogala, E., Leski, J.: Application of entropy and energy measure of fuzziness to processing of ecg signal. *Fuzzy Sets and Systems* 97(1), 9–18 (1998)
2. Gacek, A., Pedrycz, W.: A granular description of ecg signals. *IEEE Transactions on Biomedical Engineering* 53(10), 1972–1982 (2006)
3. Haykin, S.: Neural networks: A comprehensive foundation. Prentice Hall, Upper Saddle River (1999)
4. Henzel, N., Gacek, A., Leski, J.: Fuzzy sets time series in describe and analysis of ecg signals. *Acta Bio-Optica et Informatica Medica* 17(4), 313–316 (2011) (in polish)
5. Leski, J.: Neuro-fuzzy systems. WNT, Warszawa (2008) (in polish)
6. Leski, J.: Iteratively reweighted least squares classifier and its ℓ_2 - and ℓ_1 -regularized kernel versions. *Bull. Pol. Ac.: Tech.* 58(1), 171–182 (2010)
7. Moody, G., Mark, R.: The impact of the mit-bih arrhythmia database. *IEEE Eng. Med. and Biol.* 20(3), 45–50 (2001)
8. Schölkopf, B., Smola, A.: Learning with kernels. Support vector machines, regularization, optimization, and beyond. The MIT Press, London (2002)
9. Vapnik, V.: Statistical learning theory. Wiley, New York (1998)

Interpolation Procedure in Filtered Backprojection Algorithm for the Limited-Angle Tomography

Aleksander Denisiuk

Abstract. A new data completion procedure for the limited-angle tomography is proposed. The procedure is based on explicit integral interpolation formula for band-limited functions. The completed data can be used in standard filtered backprojection tomographical algorithm. The results of numerical simulation are presented.

1 Introduction

The limited-angle tomography and related problem of interpolation of band-limited function are of constant attention from early sixties till nowadays. It appears in many fields, including signal processing, medical imaging, geophysics, astronomy, electron microscopy. See [5–8, 12]. Mathematically it consists of reconstruction of a function f from its Radon transform \hat{f} known for incomplete data of angles.

The problem is characterized with strong ill-posedness. The standard regularization procedure of minimizing the L^2 norm of extention gives poor results [7]. So, different other approaches were developed for regularization, see for instance [3, 9–11]. In this article we propose a new procedure of data completion, based on explicit integral formula of interpolation of a band-limited function [1]. Data, completed with our procedure, can be used in standard filtered backprojection tomographical algorithms.

The relation of the limited-angle tomography to interpolation of band-limited functions was pointed out in [1]. However, this way was never realized in practice. The purpose of this article is to discretize the interpolation formula from [1] and to work out a procedure of completion of the tomographical data, measured with angular limitation.

To fix notations, let us introduce the standard parametrization (ω, p) on the set of lines on the plane:

University of Warmia and Mazury, Olsztyn, Poland
e-mail: denisjuk@matman.uwm.edu.pl

$$L_{\omega,p} = \{ \omega_1 x_1 + \omega_2 x_2 = p \},$$

where $\omega = (\omega_1, \omega_2) = (\cos \varphi, \sin \varphi)$ is the unit normal vector of line $L_{\omega,p}$, and p is its distance from the origin.

The Radon transform is defined as follows:

$$\hat{f}(\omega, p) = \int_{L_{\omega,p}} f(x) ds = \int f(x_1^0 + \sin \varphi t, x_2^0 - \cos \varphi t) dt, \quad (1)$$

where $x^0 = (x_1^0, x_2^0)$ is a point on a line $L_{\omega,p}$, $x_1^0 \omega_1 + x_2^0 \omega_2 = p$. One can see that $\hat{f}(-\omega, -t) = -\hat{f}(\omega, t)$, so we restrict the domain of \hat{f} to $\varphi \in [0, \pi]$.

The *limited angle problem* consists of reconstruction of a function $f(x)$ from its Radon transform $\hat{f}(\omega, p)$, given only for $\varphi \in [0, \frac{1}{2}(\pi - \varphi_0)] \cup [\frac{1}{2}(\pi + \varphi_0), \pi]$, v. Fig. 1, where cone of unknown normals is colored in gray.

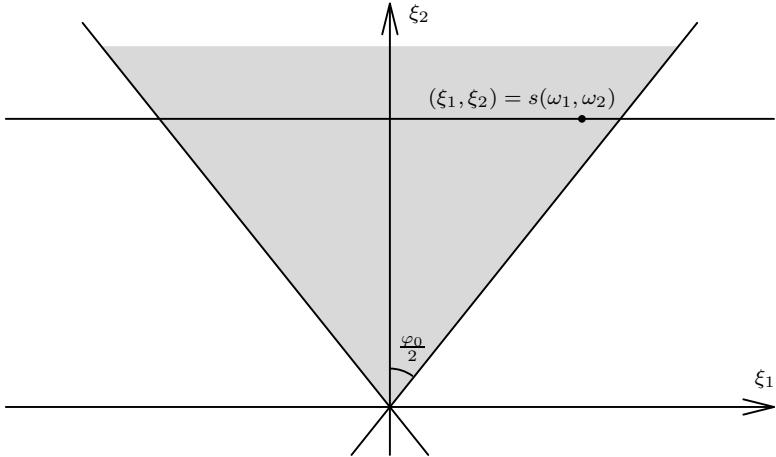


Fig. 1 Limited-angle problem. Unknown normals cone is shown in gray.

We will use the following form of the Fourier transform and its inverse:

$$\begin{aligned} \tilde{f}(\xi) &= (\sqrt{2\pi})^{-n} \int e^{-ix\xi} f(x) dx, \\ f(x) &= (\sqrt{2\pi})^{-n} \int e^{ix\xi} \tilde{f}(\xi) d\xi, \end{aligned}$$

where $x\xi = x_1 \xi_1 + \dots + x_n \xi_n$. Integration is performed over \mathbb{R}^n . We will consider only two-dimensional case, $n = 2$, but our results can be obviously generalized to higher dimensions.

Let us also recall the Fourier slice theorem ([2]):

$$\tilde{f}(s\omega) = (\sqrt{2\pi})^{-n} \int e^{-isp} \hat{f}(\omega, q) dq, \quad (2)$$

$$\hat{f}(\omega, p) = (\sqrt{2\pi})^{n-2} \int e^{isp} \tilde{f}(s\omega) ds, \quad (3)$$

We will assume that function f is supported into the unit ball, $\text{supp } f \subset \mathbb{B} = \{\|x\|^2 \leq 1\}$.

The paper consists of four sections. In section 2 we obtain explicit analytical solution of the problem, based on the integral interpolation formula from [1] (theorem 2). In section 3 we apply this formula for reconstruction of the filtered Radon transform of f , the main result of this section is the theorem 3. In section 4 we discretize the obtained formula for using in the standard filtered backprojection reconstruction algorithm. The theorem 4 is a preliminary step to discretization, it concretizes the theorem 3 to the case of filter (8). Results of the numerical simulation are presented in section 5. We use a phantom similar to that of [7] for our simulations.

2 Analytical Interpolation

Let us remind the theorem of integral interpolation of a band-limited function from [1].

Theorem 1 ([1]). Let $\psi \in L^2(\mathbb{R})$, $\text{supp } \tilde{\psi} \subset [-1, 1]$. Then for any $\rho > 0$, $\xi \in (-\rho, \rho)$

$$\psi(\xi) = \frac{1}{\pi} e^{\sqrt{\rho^2 - \xi^2}} \int_{|\eta| > \rho} \frac{\sin \sqrt{\eta^2 - \rho^2}}{|\eta - \xi|} \psi(\eta) d\eta \quad (4)$$

Since function f is supported in the unit ball, the theorem 4 is applicable to $\psi(\xi) = \tilde{f}(\xi, \xi_2)$, where ξ_2 is fixed, i.e. \tilde{f} is restricted onto the line $\xi_2 = \text{const}$, Fig. 1. In that way it is possible to interpolate the Fourier transform to the whole plane. As a result one obtains the following theorem.

Theorem 2. Let $f \in L^2(\mathbb{R}^2)$, $\text{supp } f \subset \mathbb{B}$. Then for any $\omega = (\cos \varphi, \sin \varphi)$, where $|\frac{\pi}{2} - \varphi| < \frac{\varphi_0}{2}$, the following interpolation formula is valid:

$$\hat{f}(\omega_1, \omega_2, p) = \frac{1}{2\pi^2} \int ds \int_{|\tau| > a} d\tau \int dq K_{\omega, p}(s, \tau, q) \hat{f}\left(\frac{\tau}{\sqrt{\tau^2 + \omega_2^2}}, \frac{\omega_2}{\sqrt{\tau^2 + \omega_2^2}}, q\right), \quad (5)$$

where $\omega = (\cos \varphi, \sin \varphi)$,

$$K_{\omega, p}(s, \tau, q) = \frac{\sin(s\sqrt{\tau^2 - a^2})}{|\tau - \omega_1|} e^{is(p - q\sqrt{\tau^2 + \omega_2^2}) + |s|\sqrt{a^2 - \omega_1^2}},$$

$a = \omega_2 \tan(\varphi_0/2)$, φ_0 characterizes the unknown angle range, Fig. 1

Proof. To prove the theorem, one should start with the formula (3). Then substitute into it the interpolation formula (4). And, in turn, substitute the slice theorem formula (2) into the result. Finally, one should switch from η to τ coordinate with the following substitution $\eta = s\tau$ and take use of the equality $\rho = as$.

3 Filtration

The standard tomographical reconstruction algorithm consists of two steps: filtration and backprojection. We will interpolate the Radon transform of filtered function f_b , where

$$\tilde{f}_b(\xi) = \Phi_b(|\xi|)\tilde{f}(\xi),$$

parameter b characterizes resolution of reconstruction, [7]. Here $\Phi_b(s)$ is a filter function supported in $[-b, b]$.

Filtration implies that integration with respect to s in 5 is restricted only to finite interval $[-b, b]$, so the formula 5 will be transformed as follows:

Theorem 3. Let $f \in L^2(\mathbb{R}^2)$, $\text{supp } f \in \mathbb{B}$. Then for any $\omega = (\cos \varphi, \sin \varphi)$ where $|\frac{\pi}{2} - \varphi| < \frac{\varphi_0}{2}$ the following interpolation formula for the filtered Radon transform is valid:

$$\hat{f}_b(\omega, p) = \frac{1}{\pi^2} \int_{|\tau|>a} d\tau \int dq K_{b,\omega,p}(\tau, q) \hat{f}\left(\frac{\tau}{\sqrt{\tau^2 + \omega_2^2}}, \frac{\omega_2}{\sqrt{\tau^2 + \omega_2^2}}, q\right), \quad (6)$$

where $\omega = (\cos \varphi, \sin \varphi)$,

$$K_{b,\omega,p}(\tau, q) = \int_0^b \Phi_b(s) e^{s\sqrt{a^2 - \omega_1^2}} \cos s \left(p - q\sqrt{\tau^2 + \omega_2^2} \right) \frac{\sin(s\sqrt{\tau^2 - a^2})}{|\tau - \omega_1|} ds, \quad (7)$$

$a = \omega_2 \tan(\varphi_0/2)$, φ_0 characterizes the unknown angle range, Fig. 1

4 Discrete Interpolation

Now we will discretize the formula 6. Consider the standard parallel scanning scheme [7]. It means that the following data are given for $j = 0, \dots, P-1$, $l = -Q, \dots, Q$:

$$g_{j,l} = \hat{f}(\cos \theta_j, \sin \theta_j, s_l), \text{ where } \theta_j = \pi j / P, s_l = hl, h = 1/Q.$$

Limited-angle restriction means that the data are given only for θ_j , satisfying $|\pi/2 - \theta_j| > \varphi_0/2$, i.e. $j < P\frac{\pi - \varphi_0}{2\pi}$ and, $j > P\frac{\pi + \varphi_0}{2\pi}$.

Consider the following filter, parametrized by $\varepsilon \in [0, 1]$ (cf. [7]):

$$\Phi_b(s) = \begin{cases} 1 - \varepsilon \frac{s}{b}, & 0 \leq s \leq b, \\ 0, & s > b. \end{cases} \quad (8)$$

Integral in 7 can be calculated explicitly. So, the theorem 3 can be reformulated in the following way:

Theorem 4. *Let $f \in L^2(\mathbb{R}^2)$, $\text{supp } f \in \mathbb{B}$, filtration function be defined in (8). Then for any $\omega = (\cos \varphi, \sin \varphi)$ where $|\frac{\pi}{2} - \varphi| < \frac{\varphi_0}{2}$ the following interpolation formula for the filtered Radon transform is valid:*

$$\hat{f}_b(\omega, p) = \frac{1}{2\pi^2} \int_{|\tau|>a} d\tau \int \frac{dq K_{b,\omega,p}(\tau, q)}{|\tau - \omega_1|} \hat{f}\left(\frac{\tau}{\sqrt{\tau^2 + \omega_2^2}}, \frac{\omega_2}{\sqrt{\tau^2 + \omega_2^2}}, q\right), \quad (9)$$

where $\omega = (\cos \varphi, \sin \varphi)$,

$$K_{b,\omega,p}(\tau, q) = (1 - \varepsilon) E_1(\alpha, \beta, \gamma) + \frac{\varepsilon}{b} E_2(\alpha, \beta, \gamma) + \frac{\sin v_+}{\sqrt{\alpha^2 + (\beta + \gamma)^2}} - \frac{\sin v_-}{\sqrt{\alpha^2 + (\beta - \gamma)^2}}, \quad (10)$$

where

$$E_\lambda(\alpha, \beta, \gamma) = \frac{e^{\alpha b} \sin((\beta + \gamma)b - \lambda v_+)}{(\sqrt{\alpha^2 + (\beta + \gamma)^2})^\lambda} - \frac{e^{\alpha b} \sin((\beta - \gamma)b - \lambda v_-)}{(\sqrt{\alpha^2 + (\beta - \gamma)^2})^\lambda},$$

$\alpha = \sqrt{a^2 - \omega_1^2}$, $\beta = p - q\sqrt{\tau^2 + \omega_2^2}$, $\gamma = \sqrt{\tau^2 - a^2}$, $v_\pm = \arctan((\beta \pm \gamma)/\alpha)$, $a = \omega_2 \tan(\varphi_0/2)$, φ_0 characterizes the unknown angle range, Fig. 1.

Remark 1. The case $\varepsilon = 0$ was considered in [1]. The present result has different form, more suitable for discretization.

Passing to discretization of this formula, let us make a substitution $\tau = \omega_2 \cot \theta$ in the formula (9). It will be rewritten in the following way, which is opportune for direct discretization, using given data:

$$\hat{f}_b(\omega, p) = \frac{1}{2\pi^2} \int_{|\frac{\pi}{2} - \theta| > \frac{\varphi_0}{2}} \int \frac{K_{b,\omega,p}(\omega_2 \cot \theta, q) dq d\theta}{\sin \theta |\omega_2 \cos \theta - \omega_1 \sin \theta|} \hat{f}(\cos \theta, \sin \theta, q), \quad (11)$$

Note that the standard filtered backprojection algorithm reconstructs a function of essential band width b . So, the integral with respect to η in (4) should be restricted to

finite segment $|\eta| < b$. As a corollary, the integral with respect to τ in (9) should also be restricted to some finite segment. We use the segment $|\tau| < 1.5$. This threshold was found empirically.

Finally, the filtration kernel (10) has a large multiplier $e^{\alpha b}$, which increase computational errors. This multiplier depends on the direction ω . So, the filtration level could be chosen interdependently for each direction. We use $b = \ln(50)/\alpha$, so errors are increased up to 50 times..

Summarizing, we get the following modification of the standard filtered back-projection algorithm (cf. [7]):

1. For $j = 0, \dots, P - 1$ compute $\psi_j = \frac{\pi j}{P}$
- a. if direction φ_j is given ($|\frac{\pi}{2} - \psi_j| > \frac{\varphi_0}{2}$), compute convolutions

$$v_{j,k} = 1/Q \sum_{l=-Q}^Q w_b(s_k - s_l) \hat{f}(\cos \psi_j, \sin \psi_j, s_l), \quad k = -Q, \dots, Q$$

where $w_b(s)$ is the filtration kernel corresponding to (7). Explicit expression of w_b can be found in [7].

- b. else (direction is missing) compute interpolated convolutions

$$v_{j,k} = \frac{1}{2\pi QP} \sum_{i=0}^{P-1} \sum_{l=-Q}^Q \frac{K_{b,\omega,p_k}(\omega_2 \cot \theta_i, s_l)}{\sin \theta_i |\omega_2 \cos \theta_i - \omega_1 \sin \theta_i|} \hat{f}(\cos \theta_i, \sin \theta_i, s_l),$$

where $k = -Q, \dots, Q$, $\omega = (\cos \psi_j, \sin \psi_j)$, kernel K is given in (10), summation with respect to i is performed only for those i , that satisfy $1.5\omega_2 < |\tan \theta_i|$ and $|\frac{\pi}{2} - \theta_i| > \frac{\varphi_0}{2}$.

2. At each point $x = (x_1, x_2)$ of reconstruction interpolated discrete backprojections are computed:

$$f_{bi}(x) = \frac{2\pi}{P} \sum_{j=0}^{P-1} ((1-u)v_{j,k} + uv_{j,k+1}),$$

where u and k for each pair of x and j are defined by the following formulas:

$$s = x_1 \cos \theta_j + x_2 \sin \theta_j, k \leq sQ < k + 1, u = sQ - k.$$

Remark 2. The proposed algorithm can be used with any other filter $F_b(\xi)$ (such that $F_b(\xi) = 0$ for $|\xi| > b$). If we set in (7) $\varepsilon = 0$, composition of filtrations with (7) and with F_b is equivalent to filtration only with F_b . So, discrete interpolation should be followed by discrete filtration with F_b .

5 Numerical Simulation

The results of numerical simulation are presented at the figure 2. We use a phantom similar to that of [7]. The original function has value 2 inside the ellipsoidal ring and 0.8 inside the smaller ellipse. The value of ε in filter (8) is equal to 1. There were 72 directions and 227 lines in each direction used.

In limited-angle problem we supposed that interval of 30° in unknown, i.e. 14 directional data are missing.

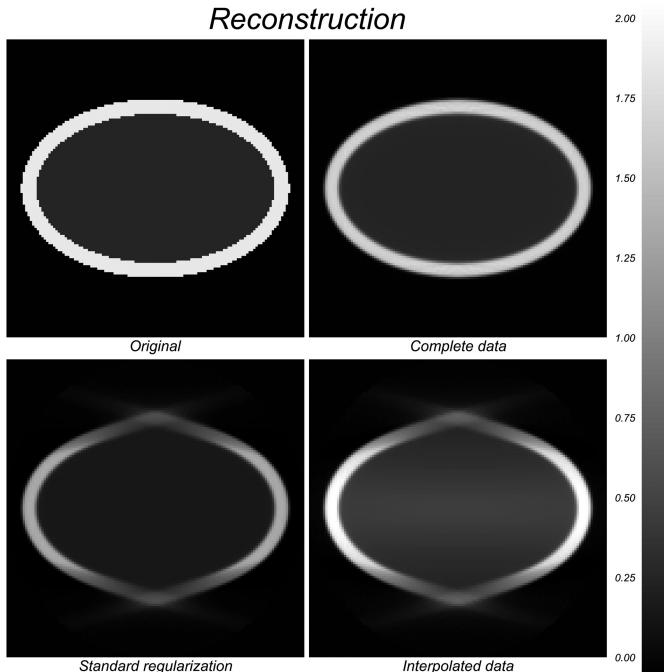


Fig. 2 Original function (top-left) and reconstruction: complete data (top-right), standard regularization procedure (bottom-left) an interpolated data (bottom-right). Unknown interval is 30° . The original function has value 2 inside the ellipsoidal ring and 0.8 inside the smaller ellipse.

Comparing errors of reconstruction, given at the figure 3, one can see significant improvement of reconstruction inside the object. At the same time outer artifact where weakened only slightly. Note that from tomographical point of view the interior of the object is the matter of interest.

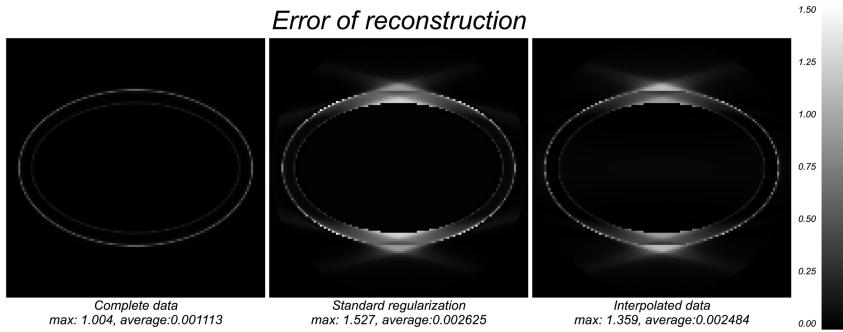


Fig. 3 Error of reconstructions (left-to-right): complete data, standard regularization procedure, interpolated data

6 Conclusions

We proposed a new universal data completion procedure for the limited-angle tomography. Our procedure can be used in the standard filtered backprojection reconstruction algorithm. The numerical simulation demonstrates that data completion allows to decrease some artifacts and improve reconstruction.

References

1. Denisjuk, A., Palamodov, V.: CRAS Paris 307, Série I:181–183 (1988)
2. Helgason, S.: The Radon Transform. Birkhauser, Boston (1999)
3. Jang, K.E., Lee, J., Lee, K., Sung, Y., Lee, S.: Proc. SPIE, vol. 8313, 83134Y (2012)
4. Kolehmainen, V., Lassas, M., Siltanen, S.: SIAM Journal on Scientific Computing 30(31413-1429) (2008)
5. Iu, K.A., Kheifets, V.N.: Kosmicheskie Issledovaniia 25, 884–894 (1987)
6. Landau, H.J., Slepian, D.D.: Bell Sys. Tech. J. 40:75, 43–63 (1961)
7. Natterer, F.: The mathematics of computerized tomography. SIAM (2001)
8. Quinto, E.T.: Inverse Problems 14, 339–353 (1998)
9. Quinto, E.T., Öktem, O.: Proc. Appl. Math. and Mech. 7, 105031–105032
10. Rantala, M., Vänskä, S., Järvenpää, S., Kalke, M., Lassas, M., Moberg, J., Siltanen, S.: IEEE Transactions on Medical Imaging 25(2), 210–217 (2006)
11. Ritschl, L., Bergner, F., Kachelriess, M.: Proc SPIE 7622, 76222H (2010)
12. Ruff, L.J.: Geophys Res. Lett. 11(7), 629–632 (1984)

Classification of Uterine Electrical Activity Patterns for Early Detection of Preterm Birth

Janusz Jezewski, Adam Matonia, Robert Czabanski,
Krzysztof Horoba, and Tomasz Kupka

Abstract. Preterm birth is the leading cause of a neonatal death, so it is extremely important to distinguish the pregnancy at risk of preterm threatening labour. Monitoring an electrical activity of the uterine muscle seems very promising as a method which enables noninvasive recording of good quality electrohysterographic signals. The developed instrumentation enabled recording of signals by means of electrodes attached to abdominal wall and determination of quantitative parameters describing the contractions detected. Our research material comprised 3 groups of patients: with physiological pregnancy, with the symptoms of premature threatening labour and patients during at term labour. The classification of uterine activity signals in each pair of groups was made using nonlinear Lagrangian Support Vector Machines which allows for improving the computational efficiency and learning quality of the SVM algorithm. The obtained results show that the proposed approach is able to differentiate between the contractile activity in physiological pregnancy and that connected with a risk of premature labour. Identification of these high-risk pregnancies leads to an enhanced perinatal surveillance.

1 Introduction

Preterm birth is the leading cause of a neonatal death. Therefore, it is extremely important to recognize a high-risk pregnant women with premature uterine contraction activity [17] [18] [22]. As it has been confirmed in clinical practice, the classical

Janusz Jezewski · Adam Matonia · Krzysztof Horoba · Tomasz Kupka
Biomedical Signal Processing, Institute of Medical Technology and Equipment,
Zabrze, Poland
e-mail: jezewski@itam.zabrze.pl

Robert Czabanski
Institute of Electronics, Silesian University of Technology, Gliwice, Poland
e-mail: robert.czabanski@polsl.pl

method based on analysis of mechanical uterine activity is not sufficient for precise discrimination of patients at risk of premature labour. Thus, it seems very promising to introduce into clinical practice the electrohysterography (EHG), as a method which enables noninvasive recording of a good quality signal of the uterine electrical activity [8] [10] [12].

Contraction of the uterine muscle cell is a result of the flow of ion currents [8] [20] [25]. In the signal being a difference between potentials recorded from two electrodes on maternal abdominal wall, the electrical activity during uterine contractions is manifested by the bursts of action potentials. The bursts occur synchronously with the mechanical uterine activity [9]. Slow wave component represents occurrence of bursts, whereas the fast wave component (with useful frequency band $0.1 \div 3$ Hz) is supposed to comprise information on electrophysiological properties of the uterine muscle. The EHG signal can be modeled as an action potentials fast wave whose amplitude is modulated by the slow wave corresponding to the contractions frequency [12] [25]. Such model allows us for a detection of the uterine contraction patterns in the electrohysterogram and their time domain analysis in a similar way as for conventional mechanical activity signal. The additional spectral parameters can be obtained exclusively from the analysis of the EHG fast wave [11] [26] [27]. Different procedures of qualitative assessment of electrohysterographic signals are proposed in literature [16] [23] [24]. Recent investigations have focused mainly on methods for predicting preterm labour on the early stage of pregnancy [21].

In this paper we described an attempt for evaluation of a risk of premature labour by classification of the EHG signals using the modified algorithm of the support vector machine (SVM) to analyze data and to recognize patterns. The machine learning methods have been successfully applied in solving a broad range of problems in perinatal medicine [13]. The early solutions [14] [15] focused mainly on the use of artificial neural networks. Recently, the learning procedures using principles of the statistical learning theory are becoming increasingly popular as powerful methods for data classification [6] [7]. In our investigation we applied the Lagrangian support vector machine (LSVM) [19], which improves the computational efficiency of the standard SVM algorithm by replacing a quadratic programming procedure of the Lagrange multipliers estimation with the linearly convergent iterative algorithm.

2 Methods

The system for acquisition and analysis of signals recorded on maternal abdomen consists of a microcontroller-based signal recorder and external computer. Typical configuration of the abdominal electrodes comprises four electrodes placed in vertical line crossing the navel and the reference electrode placed above the pubic symphysis. Additionally on the left leg, the common mode reference electrode is placed. Such configuration provides four unipolar signals to be recorded during the monitoring session (Fig. 1).

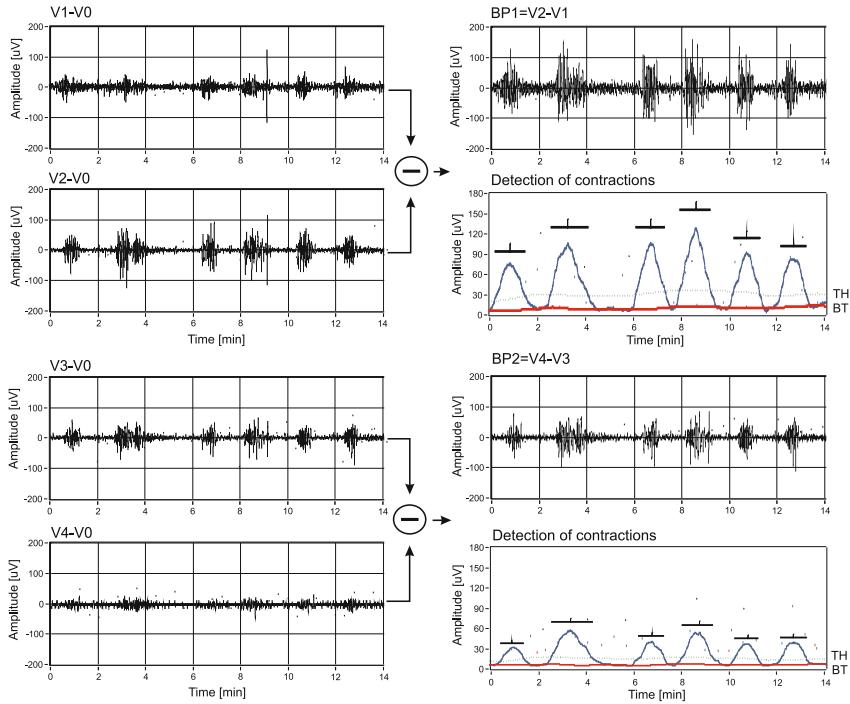


Fig. 1 Unipolar raw EHG signals (left side) relating to activity of the upper and lower part of uterine muscle and corresponding bipolar signals (right side) together with associated contraction curves and markers of contraction patterns detected.

The duration of monitoring was between 10 and 60 minutes, which allowed us to recognize from 15 to 25 uterine contractions in each signal. The patients were divided into three groups: group I with 27 patients with physiological pregnancy, group II comprising 21 patients with the symptoms of premature threatening labour, and finally group III with 14 patients during at term labour (in the first stage) [27]. Each EHG signal was represented by classical timing parameters as well as by the parameters relating to pattern description in a frequency domain [12] [26].

Recording circuit in a form of external optically isolated module enables acquisition of uterine electrical activity signals from a maternal abdomen [11]. Entire circuit enables the amplification of recorded signals from the tens of microvolts up to a few volts level. Moreover, the band-pass filtering with a lower cut-off frequency being changed from 0.05 Hz to 0.5 Hz also secure the circuit against too large low-frequency interferences. Tuning can be done using virtual instrumentation screen at the beginning of monitoring by a visual assessment of the abdominal signals recorded. The high cut-off frequency is established at 100 Hz, hence at the

sampling frequency of 500 Hz, the recorder circuit is fully protected against a possibility of aliasing occurrence.

2.1 Signal Analysis

The EHG signals underwent the off-line analysis using the LabView graphical programming environment (National Instruments). On the first stage two bipolar signals are calculated using the four unipolar channels as relating to the upper ($BP_1 = V_2 - V_1$) and the lower ($BP_2 = V_4 - V_3$) part of the uterine muscle (Fig. 1). Bipolar acquisition channels thanks to high value of CMRR are able to suppress so called common interferences and thus provide good quality of the EHG signals. However, a use of bipolar channel limits the area of uterine muscle from which the working potentials can be measured to the depth corresponding to a distance between the electrodes.

The analysis of contraction curve, that is represented in classical approach by tocogram, relies on determination of quantitative parameters describing the detected uterine contraction patterns in the time domain: occurrence rate R , duration T_D , rise time T_A , amplitude A and area S (Fig. 2). Contractions detection relies on finding those segments in the contraction curve whose amplitude exceeds the established threshold level for > 30 s. The threshold level is higher than the so called basal tone, representing some resting electrical activity of the uterine muscle cells.

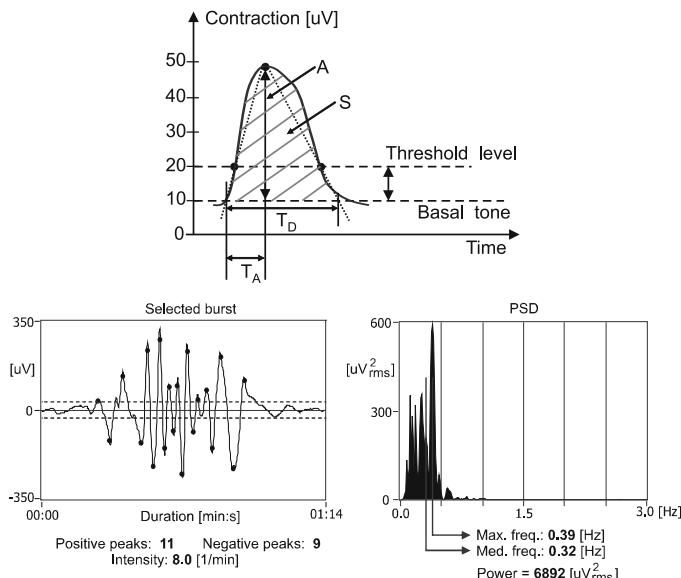


Fig. 2 Time domain parameters important for the contraction detection. Below is a segment with burst of spikes selected from raw EHG signal to determine the frequency domain parameters.

In order to detect the contractions in electrohysterogram and then to perform their classical time domain analysis, at first it is necessary to extract the slow wave corresponding to contraction curve from the acquired signal. The method of extraction of EHG slow wave is based on calculation of consecutive root-mean-square values in one-minute window stepped with 3 s. Unlike the uterine mechanical activity, where the signal values change in an established range (0 to 100 arbitrary units), the amplitude of electrohysterogram strongly depends on a given patient [12]. Therefore, the algorithm applied for contractions detection should be able to compensate the amplitude variation between electrohysterograms [12].

In every window of four-minute width and with one-minute step, samples of the slow wave are ordered from the lowest to the highest value, and then the mean value from the 10% of samples from the lower side is calculated and taken as the consecutive sample of the basal tone. So, this process can be called as the joint median and moving average filtration. The basal tone is resampled to 20 Hz. Such window comprises prelabour contraction of about 1.5 min duration together with a non-activity segment. In the same window the value of 25% of the difference between the maximum and minimum values of the slow wave is calculated and added to the basal tone samples. In that way the threshold level is determined. The contraction is recognized when the EHG slow wave remains above the threshold level for at least 30 s, and its amplitude exceeds the double distance between the basal tone and threshold level. After the contractions were detected in both bipolar signals, the signal to noise ratio (SNR) was calculated as the signal power within contractions to the segments between them. The bipolar signal of higher SNR was chosen for further analysis.

After calculation of contraction timing parameters, the segments with bursts of action potentials are selected from the raw EHG signal (Fig. 2). Contraction intensity represents a number of spikes within the burst corresponding to the uterine contraction. Action potential spike has a form of biphasic wave, which consists of two deflections in negative and positive direction. The average value of positive and negative peaks recalculated into one-minute interval is assumed as the intensity of the particular contraction. Within the burst segments the parameters exclusive for uterine electrical activity are determined: the signal power P , median frequency F_{med} and maximum power frequency F_{max} . Median frequency represents the component frequency which splits the power density spectrum into two parts comprising the same power. Maximum power frequency indicates the dominant component in power density spectrum [26].

2.2 Classification Method

The support vector machine is a learning method based on the structural risk minimization principle, which results in high learning efficiency with improved generalization capability [4]. The SVM algorithm is widely used to solve various problems relating both to classification and nonlinear regression [2] [3]. The task of the SVM classifier is to find a hyperplane in the multi-dimensional feature space,

which separates the considered classes with the largest distance (margin). The data (input vectors) that ensure the margin maximization are called the support vectors. The standard SVM procedure formulates a binary linear classifier. However, in most practical applications the considered data are not linearly separable. Thus, the input vectors are usually mapped into a space of higher dimension (using so called kernel functions), where the probability of finding the hyperplane allowing for linear separation is much higher [5].

In practical implementation a computational complexity and thus the learning speed of the classifier is crucial. Thus, we applied a modified SVM method – the Lagrangian support vector machines (LSVM), which solves the constraint minimization problem with linearly convergent iterative algorithm instead of a quadratic programming procedure. It allows improving the computational efficiency while maintaining the learning quality.

Consider a training set $\{\mathbf{x}_n, y_n\}$, consisting of N , t -dimensional input feature vectors $\mathbf{x}_n \in \mathbf{R}^t$ and corresponding class labels $y_n \in \{-1, +1\}$. With the assumed degree of misclassification $\xi \in \mathbf{R}^N$, the problem of linear SVM consists in finding parameters $\mathbf{w} \in \mathbf{R}^N$, $b \in \mathbf{R}$ of the separating hyperplane that maximizes the margin between considered classes:

$$y_n(\mathbf{w}^T \mathbf{x}_n - b) \geq 1 - \xi_n. \quad (1)$$

In order to obtain the unknown parameters the following minimization problem is defined:

$$\min_{\mathbf{w}, \xi} \frac{\mathbf{w}^T \mathbf{w}}{2} + \gamma \mathbf{e}^T \xi, \quad (2)$$

with constraints:

$$\mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}b) + \xi \geq \mathbf{e}, \quad \text{and} \quad \xi \geq 0, \quad (3)$$

where: $\mathbf{e} \in \mathbf{R}^N$ is the vector of ones, $\mathbf{D} \in \mathbf{R}^{N \times N}$ is a diagonal matrix of class labels and $\mathbf{A} \in \mathbf{R}^{N \times N}$ is the matrix of the input vectors $\mathbf{A} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T]^T$.

The LSVM replaces the sum of slack variables $\mathbf{e}^T \xi$ with the sum of their squares $\xi^T \xi$ and includes the bias b^2 in the minimization criterion:

$$\min_{\mathbf{w}, \xi, b} \frac{1}{2} (\mathbf{w}^T \mathbf{w} + b^2) + \frac{\gamma}{2} \xi^T \xi. \quad (4)$$

Its solution, under the constrain (3), is formulated as the unconstrained minimization of the Lagrangian [19]:

$$L(\lambda) = \frac{\lambda^T}{2} \left(\frac{\mathbf{I}}{\gamma} + \mathbf{D}(\mathbf{A}\mathbf{A}^T + \mathbf{e}\mathbf{e}^T)\mathbf{D} \right) \lambda - \mathbf{e}^T \lambda, \quad (5)$$

where: $\lambda \in \mathbf{R}^N$, $\lambda > 0$ are Lagrange multipliers and $\mathbf{I} \in \mathbf{R}^{N \times N}$ is the identity matrix. The classifier parameters can be obtained by partially differentiating L with respect to \mathbf{w} , b , ξ and equating the resulting equations to zero, while the unknown Lagrange multipliers are the result of the simple iterative scheme:

$$\lambda(k+1) = \mathbf{Q}^{-1}(\mathbf{e} + (\mathbf{Q}\lambda(k) - \mathbf{e} - \alpha\lambda(k))_+), \quad (6)$$

where: k is the iteration number, $\mathbf{Q} = \frac{1}{\gamma} + \mathbf{H}\mathbf{H}^T$, $\mathbf{H} = \mathbf{D}[\mathbf{A}^T - \mathbf{e}]$ and $(x)_+ = \max(0, x)$. The iterations are convergent for any $\lambda(0)$ if:

$$0 < \alpha < \frac{2}{\gamma}. \quad (7)$$

Finally, the linear LSVM classifier is defined as:

$$y(\mathbf{x}) = \text{sgn}(\lambda^T \mathbf{D}\mathbf{A}\mathbf{x} + b). \quad (8)$$

After introducing kernel functions $K(\mathbf{A}, \mathbf{x})$, the nonlinear solution can be obtained:

$$y(\mathbf{x}) = \text{sgn}(\lambda^T \mathbf{D}K(\mathbf{A}, \mathbf{x}) + b). \quad (9)$$

We used radial kernel functions in our investigations.

As inputs to the LSVM we used the quantitative parameters of uterine electrical activity patterns. The parameters of the algorithm have been established by means of the grid search method. The predictive capabilities of quantitative EHG signal parameters were evaluated with a help of receiver operating characteristic (ROC).

The area under the ROC curve (A_R) is equal to the probability that the rank of a randomly chosen instance for a risk of premature labour X is higher than a randomly chosen normal one Y [1]:

$$A_R = \text{Prob}(X > Y). \quad (10)$$

It allows us to analyze the relation between the basic measures of the prediction performance (sensitivity SE and specificity SP), as a function of the discrimination thresholds of a given EHG parameter.

3 Results and Discussion

General descriptive statistics determined for quantitative parameters of contractions detected in the slow wave component of EHG in the material collected is presented in Table 1. The diagnostic significance of all the EHG quantitative parameters was determined using the A_R values and the obtained results are listed in Table 2. We assumed that the A_R value below 0.62 relates to poor classification capabilities and thus we repeated the classification using a reduced set of inputs (only those with A_R value higher than 0.62).

That condition was fulfilled by five parameters: amplitude A , area S , signal power P as well as median and maximum power frequency F_{med} and F_{max} . Classification results as regards to each pair of the groups for a whole and limited sets of parameters are presented in Table 2. We noted an increase of classification efficiency in the all group pairs after selection of the input parameters which were characterized by higher

Table 1 Descriptive statistics of quantitative parameters of uterine contractions recognized for particular groups of patients

Parameter	Group I	Group II	Group III	Stat. significance		
				I-II	I-III	II-III
R [l/10min]	$3.2 \pm 0.6^{\&}$	3.4 ± 1.0	3.5 ± 1.1	-	-	-
T_D [s]	88.6 ± 17.6	85.1 ± 20.5	93.7 ± 25.5	-	-	-
T_A [s]	1371.2 ± 651.8	1132.7 ± 672.2	1511.5 ± 923.7	-	-	-
A [uV]	33.8 ± 53.6	62.2 ± 75.3	70.4 ± 114.2	*	*	-
S [a.u.]	1815 ± 3011	2841 ± 2809	3812 ± 5997	*	*	-
I [l/1min]	17.5 ± 3.7	18.0 ± 4.5	18.2 ± 4.2	-	-	-
P [uV ²]	1823 ± 5390	7493 ± 18098	9429 ± 28049	*	*	-
F_{med} [Hz]	0.30 ± 0.12	0.35 ± 0.10	0.33 ± 0.08	*	-	-
F_{max} [Hz]	0.25 ± 1.18	0.32 ± 0.29	0.27 ± 0.09	-	-	-

*difference between two groups which is statistically significant ($p < 0.05$).&mean \pm SD.

A_R value. Considering a high efficiency of EHG signals classification (84.21%) obtained for a pair of group I and II, we can conclude that the method proposed is able to differentiate between contractile activity in physiological pregnancy from that being symptoms of premature threatening labour. Additionally, the classification results reported for the group II and III show very similar properties of the contractions being detected during at term labour and contractions connected with a risk of premature labour.

Table 2 The A_R values for all inputs (contraction parameters) obtained for particular group pairs together with classification efficiency of EHG signals relating to original and reduced set of parameters

Parameter	Group I-II	Group I-III	Group II-III
R [l/10min]	0.539	0.522	0.517
T_D [s]	0.550	0.585	0.622
T_A [s]	0.612	0.524	0.639
A [uV]	0.690	0.675	0.520
S [a.u.]	0.690	0.677	0.507
I [l/1min]	0.541	0.587	0.546
P [uV ²]	0.677	0.680	0.527
F_{med} [Hz]	0.700	0.659	0.531
F_{max} [Hz]	0.628	0.634	0.514
Set of parameters	Classification efficiency		
Original	57.89%	64.71%	47.06%
Reduced ($A_R > 0.62$)	84.21%	82.35%	57.14%

Our study has shown that a set of parameters can be obtained by analysis of the uterine electrical activity signals both in time and frequency domain. These parameters can be used as inputs for the modern classification methods. The obtained results are similar to those obtained in our previous study [27] where a simple statistical approach was applied to find the difference between the groups. However, the LSVM-based classification of the EHG signals has an advantage over the simple statistical approach. This provides a clinically useful tool that can classify signals during on-line monitoring and thus inform on symptoms of premature threatening labour.

The noninvasive electrohysterography provides more complete information on functioning of the uterine muscle than classical analysis of mechanical activity so it may play a leading role in modern perinatology.

References

1. Bamber, D.: The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* 4, 387–415 (1975)
2. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2, 121–167 (1998)
3. Byun, H., Lee, S.W.: Applications of support vector machines for pattern recognition: A survey. In: Proc. of 1st International Workshop on Pattern Recognition with Support Vector Machines, vol. 2, pp. 213–236 (2002)
4. Cortes, C., Vapnik, V.: Support-vector networks. *Machine Learning* 20, 273–297 (1995)
5. Cover, T.M.: Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. on Electronic Computers* 14, 326–334 (1965)
6. Czabanski, R., Jezewski, J., Matonia, A., et al.: Computerized Analysis of Fetal Heart Rate Signals as the Predictor of Neonatal Acidemia. *Expert Syst. Appl.* 39(15), 11846–11860 (2012)
7. Czabanski, R., Jezewski, M., Wrobel, J., et al.: Predicting the Risk of Low Fetal Birth Weight from Cardiotocographic Signals using ANBLIR System with Deterministic Annealing and e-Insensitive Learning. *IEEE T. Inf. Technol. B.* 14(4), 1062–1074 (2010)
8. Devedeux, D., Marque, C., Mansour, S., et al.: Uterine Electromyography: a critical review. *Am. J. Obstet. Gynecol.* 169, 1636–1653 (1993)
9. Euliano, T.Y., Nguyen, M.T., Darmanjian, S., et al.: Monitoring uterine activity during labor: a comparison of 3 methods. *Am. J. Obstet. Gynecol.* 208, 66.e1–6 (2013)
10. Harrison, A., Crowe, J.A., Hayes-Gill, B.R., et al.: Use of the electrohysterogram for uterine contraction monitoring during labour. In: Proc. of 3rd European Conference on Engineering and Medicine, pp. 90–95 (1995)
11. Jezewski, J., Horoba, K., Matonia, A., et al.: A new approach to cardiotocographic fetal monitoring based on analysis of bioelectrical signals. In: Proc. of 25th Int. Conf. of IEEE Engineering in Medicine and Biology Society, pp. 3145–3149 (2003)
12. Jezewski, J., Horoba, K., Matonia, A., et al.: Quantitative analysis of contraction patterns in electrical activity signal of pregnant uterus as an alternative to mechanical approach. *Physiol. Meas.* 26, 753–767 (2006)

13. Jezewski, J., Wrobel, J., Horoba, K., et al.: Computerized perinatal database for retrospective qualitative assessment of cardiotocographic traces. In: Richards, B. (ed.) Current Perspectives in Healthcare Computing, pp. 187–196. BJHC Limited, Great Britain (1996)
14. Jezewski, J., Wrobel, J., Horoba, K., et al.: Fetal heart rate variability: clinical experts versus computerized system interpretation. In: Proc. of 24th Int. Conf. of IEEE Engineering in Medicine and Biology Society, pp. 1617–1618 (2002)
15. Jezewski, M., Wrobel, J., Labaj, P., et al.: Some Practical Remarks on Neural Networks Approach to Fetal Cardiotocograms Classification. In: Proc. of 29th Int. Conf. of IEEE Engineering in Medicine and Biology Society, pp. 5170–5173 (2007)
16. La Rosa, P.S., Nehorai, A., Eswaran, H., et al.: Detection of uterine MMG contractions using a multiple change point estimator and the K-means cluster algorithm. *IEEE T. Biomed. Eng.* 55(2), 453–467 (2008)
17. Lucovnik, M., Kuon, R.J., Chambliss, L.R., et al.: Use of uterine electromyography to diagnose term and preterm labor. *Acta Obstet. Gynecol. Scand.* 90(2), 150–157 (2011)
18. Maner, W., Garfield, R., Maul, H., et al.: Predicting term and preterm delivery with trans-abdominal uterine electromyography. *Obstet. Gynecol.* 101, 1254–1260 (2003)
19. Mangasarian, O., Musicant, D.: Lagrangian support vector machines. *J. Mach. Learn. Res.* 1, 161–177 (2001)
20. Rabotti, C., Mischi, M., Oei, S.G., et al.: Noninvasive estimation of the electrohysterographic action-potential conduction velocity. *IEEE T. Biomed. Eng.* 57(9), 2178–2187 (2010)
21. Takagi, K., Satoh, K., Muraoka, M., et al.: A mathematical model for predicting outcome in preterm labour. *J. Int. Med. Res.* 40(4), 1459–1466 (2012)
22. Verdenik, I., Pajntar, M., Leskosek, B.: Uterine electrical activity as predictor of preterm birth in women with preterm contractions. *Eur. J. Ostet. Gynecol. Reprod. Biol.* 95, 149–153 (2001)
23. Vrhovec, J., Rudel, D., Lebar, A.M.: The importance of uterine contractions extraction in evaluation of the progress of labour by calculating the values of sample entropy from uterine electromyogram. In: IFMBE Proceedings, vol. 29, pp. 140–143 (2010)
24. Warrick, P.A., Hamilton, E.F., Precup, D., et al.: Identification of the dynamic relationship between intrapartum uterine pressure and fetal heart rate for normal and hypoxic fetuses. *IEEE T. Biomed. Eng.* 56(6), 1587–1597 (2009)
25. Zietek, J., Sikora, J., Horoba, K., et al.: Mechanical and electrical uterine activity. Part I. Contractions monitorin. *Ginekol. Pol.* 79(11), 791–797 (2008) (in Polish)
26. Zietek, J., Sikora, J., Horoba, K., et al.: Mechanical and electrical uterine activity. Part II. Contraction parameters. *Ginekol. Pol.* 79(11), 798–804 (2008) (in Polish)
27. Zietek, J., Sikora, J., Horoba, K., et al.: Prognostic value of chosen parameters of mechanical and bioelectrical uterine activity in prediction of threatening premature labour. *Ginekol. Pol.* 80(3), 193–200 (2009) (in Polish)

Diagnosis of Bipolar Disorder Based on Principal Component Analysis and SVM

M. Termenon, Manuel Graña* A. Besga, J. Echeveste,
J.M. Pérez, and A. Gonzalez-Pinto

Abstract. In this paper, we present a computer aided diagnosis tool to discriminate between healthy subjects and patients with bipolar disorder (BD) using the deformation Jacobian obtained during the registration process of their structural (T1) magnetic resonance imaging (MRI) acquisition. To be able to compare MRI images of different subjects, first we need to register those images to a common template. We perform a two step registration: affine and nonlinear. To avoid the curse of dimensionality, we reduce the dimensions of each image performing a feature selection and linear transformation process. First, we select the voxels with the square difference above a given threshold and, second, we apply principal component analysis (PCA) for further dimensionality reduction of the selected features. Results are obtained over an on-going study in Hospital de Santiago Apostol collecting anatomical T1-weighted MRI volumes from healthy control subjects and BD patients. We perform several experiments with different thresholds achieving up to 90% of accuracy when classifying the selected features with linear support vector machines (SVM) classifier.

M. Termenon · Manuel Graña
Grupo de Inteligencia Computacional, UPV/EHU
www.ehu.es/ccwintco

A. Besga · A. Gonzalez-Pinto
Unidad de Investigación en Psiquiatría, Hospital Santiago Apostol, Vitoria-Gasteiz

J. Echeveste
Departamento de Resonancia Magnética, Osatek-Vitoria

J.M. Pérez
Servicio de Neurología, Hospital Santiago Apostol, Vitoria-Gasteiz

* Corresponding author.

1 Introduction

Bipolar disorder (BD) is a psychiatric disorder that implies at least one episode of mania or hypomania or a mixed episode which is usually related to a depressive episode, changes in mood states and psychotic symptoms [25]. It is associated with cognitive, affective and functional impairment [2]. A diagnosis of bipolar disorder is made on the basis of symptoms, course of illness and, when available, family history, but there are multiple studies based on neuroimaging that have identified several regions that are affected by the disease [1, 2, 10, 21]. Magnetic resonance imaging (MRI) obtains images based on the different water content among different structures. These images can be processed to measure volume or shape of, in our case, brain structures.

In this paper, we compare brain structural MRI of healthy controls with patients with bipolar disorder, to assess if we are able to discriminate between both groups selecting relevant information embedded in the images. We focus on the application of Machine Learning (ML) algorithms on the basis of feature vectors extracted from the deformation of the structural MRI images in order to create computer aided diagnosis (CAD) tools. CAD systems are desirable in order to improve the prediction accuracy complementing the neuropsychological assessments performed by expert clinicians [4, 7]. Previous to the feature extraction process, a careful registration of the volumes is performed, including affine and non-linear registrations to a standard template, to ensure that the same voxel site corresponds to the same anatomical location across subjects. After non-linear registration, we obtain the deformation transformation required to match the template to the corresponding subject image. The Jacobian of the deformation at each voxel will be used to extract the relevant features as done in several relevant previous works [4, 11, 17].

Previous works on ML for CAD based on MRI data have performed the feature extraction on structural MRI anatomical data using Voxel Based Morphometry (VBM) [14, 16], Pearson's correlation across volumes [2, 5, 7, 22] and other several pattern recognition techniques such as principal component analysis (PCA) [4], lattice independent component analysis (LICA) [23], partial least square (PLS) [12] and so on. In this paper, we use PCA to reduce the dimensionality of our dataset to later on, classify the selected features using a linear kernel support vector machine (SVM) classifier [3, 24].

In section 2, we present the database of the experiment, we recall the definition of PCA, and we explain the feature selection process. Next section presents the experiments and obtained results and finally, in section 4, we present the discussion and our conclusion.

2 Materials and Methods

2.1 Database

Patients included in the present study were referred to the psychiatric unit at Alava University Hospital, Vitoria (Spain) from the hospital catchment area for the

investigation of memory complaints. All patients were living independently in the community. Selected subjects underwent a standard protocol including: clinical evaluation, a cognitive and a neuropsychological evaluation, and brain imaging (MRI).

Forty men and women elderly subjects were included in the present study. The healthy control group included 20 subjects without memory complaints (mean age 74.10 (SD:8.03 years)) and BD group included 20 subjects fulfilling DSM IV's criteria [13] (mean age 70.37 (SD: 9.07 years)). Subjects with psychiatric disorders (i.e. major depression) or other conditions (i.e. brain tumors) were not considered for this study.

Structural MRI data were used to validate the proposed CAD approach. MR scanning was performed on a 1.5 Tesla scanner (Magnetom Avanto, Siemens). Study protocol consists of 3D T1-weighted acquisition (isometric 1x1x1mm, 176 slices, TR=1900ms, TE=337ms and FOV=256/76%), a 3D Flair sequence (isometric 1x1x1mm, 176 slices, TR=5000ms, TE=333ms and FOV=260/87.5%).

2.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a non-parametric method to extract relevant information from multidimensional data sets. PCA reduces the complexity of the data set projecting it to a lower dimension and revealing simplified structures that underlie it [9]. Each principal component is a linear combination of the original variables and they are orthogonal to each other.

PCA computation consists of five main steps [18]:

1. Consider that the original data set is a matrix $\mathbf{X}^{m \times N}$ of $m \times N$, whose rows indicate the features and columns the samples.
2. Subtract the mean of the data set \mathbf{X} to obtain a **zero mean dataset**: $\mathbf{Z} = \mathbf{X} - \bar{\mathbf{x}}\mathbf{1}^N$, where $\bar{\mathbf{x}} = (1/N) \sum_{i=1}^N \mathbf{x}_i$.
3. Compute the **data covariance matrix**: $\mathbf{C}^{m \times m} = \mathbb{E}[\mathbf{Z}\mathbf{Z}^T]$, where \mathbb{E} denotes the expected value.
4. Perform the eigen-decomposition of $\mathbf{C}^{m \times m}$ obtaining a matrix of **eigenvectors** $\mathbf{V} = \{\mathbf{v}_i\}$, $i = 1, \dots, m$ and a vector of **eigenvalues** $\mathbf{D} = [\lambda_1, \dots, \lambda_m]$. The eigenvectors correspond to the **axes of the principal components**.
5. **Sort** the principal components by their respective eigenvalues in decreasing order. Choose the $p < m$ first eigenvectors, \mathbf{V}_{sel} .
6. Obtain the **new data set**: $\mathbf{X}_{PCA} = \mathbf{V}_{sel}\mathbf{Z}^T$

2.3 Feature Extraction Process

Image Preprocessing

The careful preprocessing of the data is of paramount importance. After converting DICOM images into NIfTI format, T1-weighted sMRI volumes were skull stripped.

A two step registration process was then performed: first, we applied an affine registration to the Montreal Neurological Institute (MNI152) standard template; second, a non linear registration procedure using the default parameters and a scaled conjugate gradient minimization method. The spatial normalization of each subject of the database has been performed with FSL FLIRT and FSL FNIRT [19] with a template resolution of $2 \times 2 \times 2$. A visual check has been performed for all images in every processing step carried out in this experiment. In Fig. 1, we show the pipeline followed to preprocess the images.

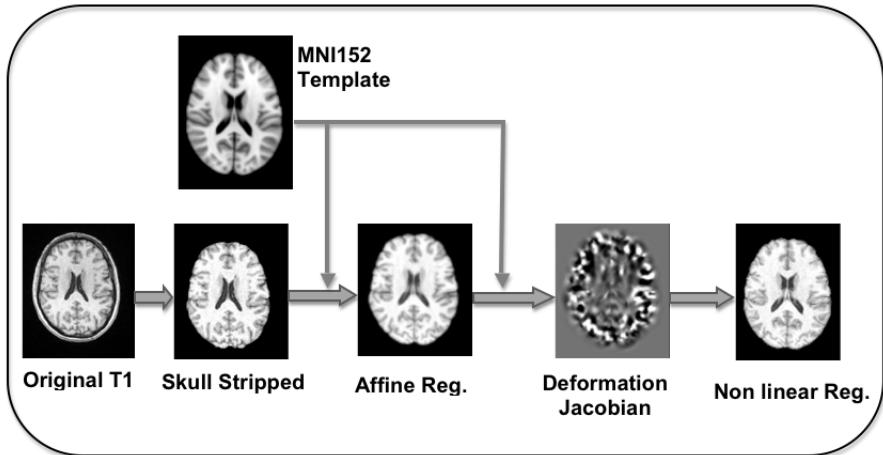


Fig. 1 Preprocessing pipeline. In this image, we show the brain of a healthy subject. First image is the original acquisition (original T1). We extract the brain to obtain the skull stripped image. Using a standard template (MNI152), we apply a linear (affine) registration and then, a non linear registration. Deformation Jacobian is the transformation needed to obtain the final non linear registered image.

Tensor Based Morphometry (TBM)

The objective of TBM is to localize regions showing shape differences among groups of brains, based on deformation fields. A template T is non-linearly registered to a subject S , obtaining a displacement vector $\vec{u}(r)$ such that $T(\vec{r} - \vec{u})$ corresponds with $S(\vec{r})$, where \vec{r} denotes the voxel location. So, the Jacobian matrix describes the local volume change of the deformation in the neighborhood of a given voxel. It is defined by:

$$\mathbf{J}_i = \begin{pmatrix} \partial(x-u_x)/\partial x & \partial(x-u_x)/\partial y & \partial(x-u_x)/\partial z \\ \partial(x-u_y)/\partial x & \partial(x-u_y)/\partial y & \partial(x-u_y)/\partial z \\ \partial(x-u_z)/\partial x & \partial(x-u_z)/\partial y & \partial(x-u_z)/\partial z \end{pmatrix}$$

In principle, the Jacobian matrices of the deformations should be more reliable indicators of local brain shape than absolute deformations [6]. In this experiment, we use the Jacobian matrices to find discriminant features for disease detection in a

supervised classification algorithm. The determinant of the Jacobian matrix, \mathbf{J}_i , is the most commonly used scalar measure of deformation for TBM analysis [11] and it is commonly used to analyze the distortion necessary to deform the images into agreement. When the change is near zero, the determinant of \mathbf{J}_i ($\det(\mathbf{J}_i)$), is also zero, meaning that there is no local difference in volume between subject and target images. If $\det(\mathbf{J}_i) > 1$ then, the neighborhood adjacent to the displacement vector in voxel i was stretched to match the template (i.e., local volumetric expansion), while if $\det(\mathbf{J}_i) < 1$, then it is associated with local shrinkage.

Discriminant Feature Selection

Voxel sites with similar values in healthy subjects and BD patients are discarded since they do not provide discriminant information. We compute the mean of the controls and the mean of the patients of the training set for each voxel site. Then, we compute the square difference between both means. The resulting volume highlights the voxels sites with maximal differences between both groups. We define a mask by selecting voxel sites with difference value above a given threshold, which is determined by inspection of the histogram of the voxels' differences distribution. That mask will be applied to the whole data set, obtaining the feature vectors that will be the input to the principal component analysis step. In Fig. 2, we show the locations of the voxels for a threshold=0.4.

PCA Transformation

In Fig. 3, we show a plot of the first three feature dimensions (left) and the three principal components obtained after PCA transformation (right). It is possible to see how, after PCA transformation, features appear spreader and it seems easier to separate between groups.

3 Experiments and Results

We perform a leave one out cross-validation technique. To quantify the results, we compute $F - score$ (also called, $F_1 score$ or $F - measure$) which is a measure of the test's accuracy. It is defined as the harmonic mean between *precision* and *recall* (also known as *sensitivity*) [15]:

$$F - score = 2 \cdot \frac{precision \cdot recall}{precision + recall},$$

where *precision* is defined as the positive predictive value, $precision = \frac{TP}{TP+FP}$, and *recall* is referred as the true positive rate, $recall = \frac{TP}{TP+FN}$. We also compute two other scores, $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ and $Specificity = \frac{TN}{TN+FP}$; where true positives (TP) are the number of patient volumes correctly classified; true negatives

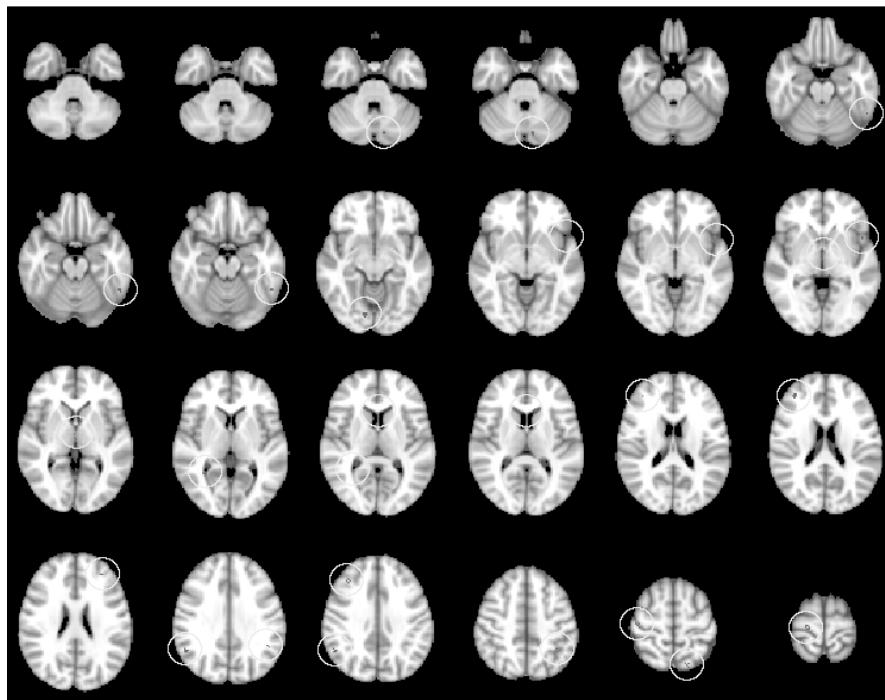


Fig. 2 Voxel location (in red) for a threshold=0.4 over a MNI152 template with voxel resolution of $2 \times 2 \times 2$. We add circles in yellow with clarification purposes.

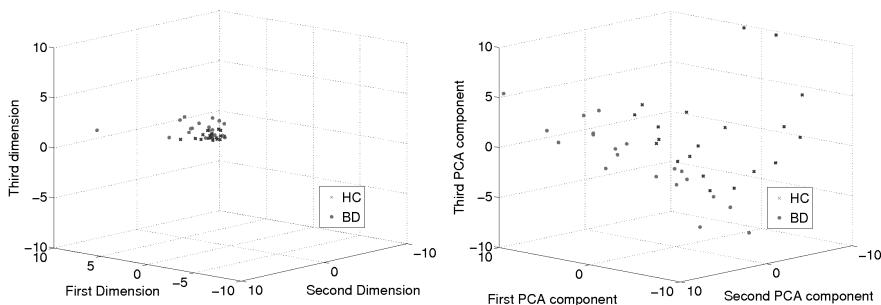


Fig. 3 Left: first three dimension of the experimental data. Right, first three dimensions of original data; left, first three PCA components of the experimental data. In blue (asterisk) healthy controls are shown. In red (filled circles) patients with bipolar disorders.

(TN) are the number of control volumes correctly classified; false positives (FP) are the number of control volumes classified as diseased patients; false negatives (FN) are the number of diseased patient volumes classified as control subjects and N , the total of subjects. We have labeled patients as class 0 and controls as class 1.

We tried to classify with different number of PCA projections. We start classifying with first PCA component, and add a new dimension each iteration. Finally, we select the dimension with the best performance in terms of F-score. In Tab. 1, we show the classification results for different thresholds applied over the square difference between the means of healthy subjects and patients. Results are encouraging, reaching 90% accuracy and 95% sensitivity.

Table 1 Classification results for a features selection procedure based on PCA and a linear SVM as classifier. Th: threshold; Acc: accuracy; Sens=sensitivity or recall; Spec: specificity. In brackets, we show the number of features selected.

Th (#ft)	0.1 (6759)		0.2 (1257)		0.3 (331)		0.4 (114)		0.5 (49)	
%	All	PCA(6)	All	PCA(8)	All	PCA(21)	All	PCA(16)	All	PCA(22)
Acc	60.00	72.50	67.50	67.50	75.00	72.50	70.00	90.00	72.50	77.50
Sens	65.00	80.00	70.00	75.00	75.00	85.00	65.00	95.00	75.00	85.00
Spec	55.00	65.00	65.00	60.00	75.00	60.00	75.00	85.00	70.00	70.00
F	61.90	74.42	68.29	69.77	75.00	75.56	68.42	90.48	73.17	79.07

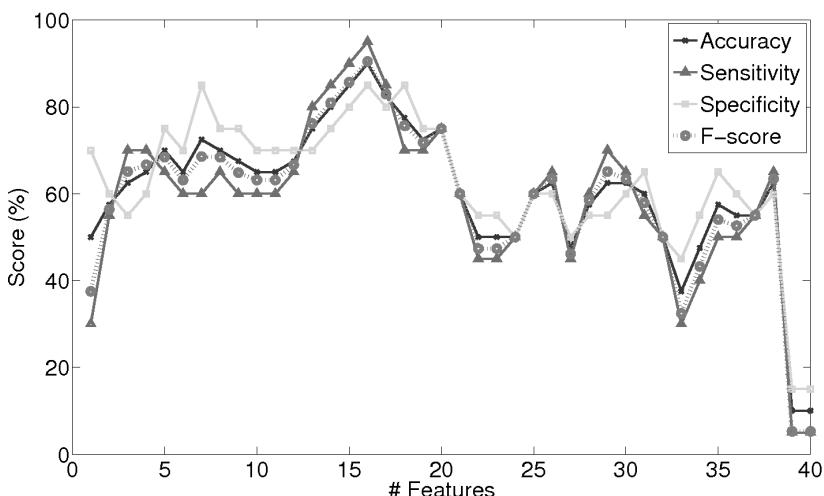


Fig. 4 Scores for different number of features when applying a threshold=0.4. We show 4 different scores: accuracy (Acc), sensitivity (Sens), specificity (Spec) and F-score.

4 Discussion and Conclusion

In this paper, we present a CAD tool to discriminate between healthy controls and BD patients using the deformation Jacobian obtained during the registration process of T1-MRI to MNI152 template. After two-step registration process, we perform a feature selection based on the square difference of voxel values between groups and PCA. Once we have selected the relevant features, we train and test a SVM classifier to discriminate between controls and BD patients.

Applied methodology is free of circularity because we perform a leave one out cross-validation technique when computing the voxels square difference between groups and, in addition, PCA is a no supervised technique. We subsample images to a $2 \times 2 \times 2$ mm resolution, considering that discriminant features of BD are not to be found at voxel resolution level but at a higher structural level. This way, we reduce the complexity without loss of information. The idea of using the deformation Jacobian instead of the final registered image is because Jacobian matrices of the deformations should be more reliable indicators of local brain shape than absolute deformations [6].

Selected voxels are mainly located in thalamus and angular gyrus, but also in precuneous cortex, precentral and postcentral gyrus, supramarginal gyrus, right lateral ventricle, superior parietal lobe, inferior temporal gyrus and cerebellum. Thalamus is one of the most relevant biomarkers in bipolar disorder [8, 10, 20] but also superior parietal lobe [1], precuneous cortex, precentral gyrus and cerebellum [8] are found in the literature as disorder biomarkers. These findings validate our approach because with no a priori information related with the disease, we find regions in the brain that are discriminant and they match with the medical literature related with bipolar disorder.

Main limitation of this study is that results come from a small database. Therefore, more extensive testing will be needed to confirm our conclusions.

References

1. Adler, C.M., Levine, A.D., DelBello, M.P., Strakowski, S.M.: Changes in gray matter volume in patients with bipolar disorder. *Biological Psychiatry* 58(2), 151–157 (2005)
2. Besga, A., Termenon, M., Graña, M., Echeveste, J., Pérez, J.M., Gonzalez-Pinto, A.: Discovering Alzheimer’s disease and bipolar disorder white matter effects building computer aided diagnostic systems on brain diffusion tensor imaging features. *Neuroscience Letters* 520(1), 71–76 (2012)
3. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2, 121–167 (1998)
4. Duchesne, S., Caroli, A., Geroldi, C., Barillot, C., Frisoni, G.B., Collins, D.L.: MRI-Based automated computer classification of probable AD versus normal controls. *IEEE Transactions on Medical Imaging* 27(4), 509–520 (2008)

5. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: COMPARE: classification of morphological patterns using adaptive regional elements. *IEEE Transactions on Medical Imaging* 26(1), 93–105 (2007) PMID: 17243588
6. Frackowiak, R.S.J.: *Human Brain Function*. Academic Press (2004)
7. Graña, M., Termenon, M., Savio, A., Gonzalez-Pinto, A., Echeveste, J., Pérez, J.M., Besga, A.: Computer aided diagnosis system for alzheimer disease using brain diffusion tensor imaging features selected by pearson’s correlation. *Neuroscience Letters* 502(3), 225–229 (2011) PMID: 21839143
8. Houenou, J., Frommberger, J., Carde, S., Glasbrenner, M., Diener, C., Leboyer, M., Wessa, M.: Neuroimaging-based markers of bipolar disorder: Evidence from two meta-analyses. *Journal of Affective Disorders* 132(3), 344–355 (2011)
9. Jolliffe, I.T.: *Principal Component Analysis*. Springer (October 2002)
10. Keener, M.T., Phillips, M.L.: Neuroimaging in bipolar disorder: A critical review of current findings. *Current Psychiatry Reports* 9(6), 512–520 (2007) PMID: 18221633 PMCID: PMC2686113
11. Lepore, N., Brun, C., Chou, Y.Y., Chiang, M.C., Dutton, R.A., Hayashi, K.M., Luders, E., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T., Thompson, P.M.: Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. *IEEE Transactions on Medical Imaging* 27(1), 129–141 (2008) PMID: 18270068
12. Nestor, P.G., O’Donnell, B.F., McCarley, R.W., Niznikiewicz, M., Barnard, J., Jen Shen, Z., Bookstein, F.L., Shenton, M.E.: A new statistical method for testing hypotheses of neuropsychological/MRI relationships in schizophrenia: partial least squares analysis. *Schizophrenia Research* 53(1-2), 57–66 (2002)
13. Ng, B., Camacho, A., Lara, D.R., Brunstein, M.G., Pinto, O.C., Akiskal, H.S.: A case series on the hypothesized connection between dementia and bipolar spectrum disorders: bipolar type VI?. *Journal of Affective Disorders* 107(1-3), 307–315 (2008) PMID: 17889374
14. Nieuwenhuis, M., van Haren, N.E.M., Hulshoff Pol, H.E., Cahn, W., Kahn, R.S., Schnack, H.G.: Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *NeuroImage* 61(3), 606–612 (2012)
15. Van Rijsbergen, C.J.: *Information retrieval*. Butterworths (1979)
16. Savio, A., Garcia-Sebastian, M.T., Chyzyk, D., Hernandez, C., Graña, M., Sistiaga, A., Lopez de Munain, A., Villanua, J.: Neurocognitive disorder detection based on feature vectors extracted from vbm analysis of structural mri. *Computers in Biology and Medicine* 41, 600–610 (2011)
17. Savio, A., Graña, M.: Supervised classification using deformation-based features for alzheimer’s disease detection on the OASIS cross-sectional database. In: *Frontiers in Artificial Intelligence and Applications*, vol. 243 (2012)
18. Smith, L.: A tutorial on principal component analysis. Technical report, Department of Computer Science, University of Otago (2002), http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
19. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobniak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M.: Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23(suppl. 1), S208–S219 (2004) PMID: 15501092
20. Strakowski, S.M., DelBello, M.P.: Brain magnetic resonance imaging of structural abnormalities in bipolar disorder. *Archives of General Psychiatry* 56(3), 254–260 (1999)

21. Strakowski, S.M., DelBello, M.P., Adler, C., Cecil, K.M., Sax, K.W.: Neuroimaging in bipolar disorder. *Bipolar Disorders* 2(3), 148–164 (2000)
22. Termenon, M., Graña, M., Barrós-Loscertales, A., Bustamante, J.C., Ávila, C.: Cocaine dependent classification using brain magnetic resonance imaging. In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B. (eds.) HAIS 2012, Part II. LNCS, vol. 7209, pp. 448–454. Springer, Heidelberg (2012)
23. Termenon, M., Graña, M., Besga, A., Echeveste, J., Gonzalez-Pinto, A.: Lattice independent component analysis feature selection on diffusion weighted imaging for alzheimer's disease classification. *Neurocomputing* (2012)
24. Vapnik, V.N.: Statistical Learning Theory. Wiley-Interscience (September 1998)
25. Vieta, E., Phillips, M.L.: Deconstructing bipolar disorder: A critical review of its diagnostic validity and a proposal for DSM-V and ICD-11. *Schizophrenia Bulletin* 33(4), 886–892 (2007)

Genetic Algorithms in EEG Feature Selection for the Classification of Movements of the Left and Right Hand

Izabela Rejer

Abstract. The crucial problem which has to be solved when an effective brain-computer interface (BCI) is to be design is: how to reduce the huge space of features extracted from raw EEG signals? One of the techniques of feature selection often used by BCI researches are genetic algorithms (GA). This approach, in its classic form, allows obtaining a feature set which gives the high classification precision, however, the dimension of this set is often still too large to create a reliable classifier. The paper presents a modified version of genetic algorithm, which is capable of choosing feature sets of a slightly lower classification precision but significantly smaller number of features. The practical application of the proposed algorithm will be presented via a benchmark EEG set submitted to the second BCI Competition (data set III - motor imaginary).

1 Introduction

Assuming the recording of EEG signals from only 32 electrodes and calculating the signal power separately in 5 frequency bands and individually per each of 5 seconds, the number of features in BCI feature space would be equal to 800. Obviously such a large number of features could not be introduced to the classifier due to large number of trials which had to be performed in order to collect enough training data. Since, it is recommended to use at least 10 times more training data per class than the features [7], at least 16000 observations would be necessary in case of only two classes. Of course, the given number of 16000 observations would be enough only in case of problems with linear or slightly non-linear characteristic [3]. In case of a strong non-linearity, the given ratio of 10 observations per class should be related to the number of the classifier free parameters rather than to the number of features, which additionally enlarges the required amount of data.

Izabela Rejer

West Pomeranian University of Technology, Szczecin; Faculty of Computer Science;

Zolnierska 49, 71-210 Szczecin

e-mail: irejer@wi.zut.edu.pl

Carrying out 16000 experiments in case of analyzing EEG signals is impossible for at least several reasons. First of all, it is the time which would be needed for dealing with this task that creates a problem. Assuming that one experiment takes 10 seconds - 16000 experiments take 45 hours. If it was possible to carry out the experiments simultaneously with 45 subject, the task would not be so hard. Unfortunately, the oscillations forming EEG signals are subject specific, which means that all data applied to the classifier has to be collected from the same subject.

The second problem is the habituation phenomenon [6]. This phenomenon means that when the experiment time extends, the power of brain rhythms extenuates. The explanation of the habituation phenomenon is that when new skills are mastered, higher cortical areas overtake the functions of the lower ones and the lower areas become less involved in the recognition process.

Apart from the habituation effect, certain aspects of subject condition such as high level of fatigue or varying level of concentration and attention make it impossible to conduct a large number of trials with the same subject. Theoretically, it should be possible to divide the whole experiment into a set of sessions, carried out over a longer period of time, but in practice it would be very difficult to preserve comparable conditions of succeeding sessions. Health problems, changes in mental condition or changes in attitude towards the experiment are factors which may change the characteristic of the subject cortical rhythms. When such situations occur during the regular application of the interface, they have a negative impact only on the temporary precision of classification; however, if they occur during the process of establishing the interface, to be exact - while training the classifier, all the future results are biased. That is why, during the process of building the interface, all sessions should be carried out in very similar conditions, that is with a stable mental and physical state of the subject, who is the most variable factor of such experiments. Hence, a more reasonable solution than to make a few days pauses between succeeding sessions is to carry out the whole experiment, divided into sessions and small breaks, on the same day.

Taking into account the aforementioned facts, it is often that the number of records collected from one subject is no more than 100-200. Such a small number of records allows building a classifier of about 5-10 input variables (assuming: 2 classes problem, 10 observations per each class and introducing the whole amount of data to the classifier training set). Comparing the huge dimension of the whole set of features extracted from raw EEG signal, often exceeding 1000, with 5-10 features which can be introduced to a classifier, it is obvious that the great reduction of the original feature space is needed.

One of the methods for feature selection, applied frequently in BCI domain, are genetic algorithms [6] [8]. Their main advantage is the fact that during the exploration of the space of possible solutions, they do not evaluate solutions one by one, but evaluate a set of solutions in parallel. Moreover, they are not very prone to get stuck at local minima and they do not need to make assumptions about the interactions between features [5].

Another feature selection method, popular in the domain of pattern recognition but rather rarely used in case of EEG feature sets, is a step-selection method.

A step-selection means that features are added/removed to/from the feature set one by one in succeeding steps of the survey. At each step this feature is added/removed to/from a feature set which is the best/worst from the point of view of a given criterion. The main drawback of this method, often underlined in scientific publications, is that it explores only one possible path in the search space, which made it prone to get stuck at local minima. This is true in general, however, when a very small subset of features is to be selected in the search process, the step-selection method (to be exact, the forward step-selection method) looking only for one feature in each of a few succeeding steps is much more capable of finding the set of features of high discrimination capabilities than genetic algorithm evaluating the huge subsets of features at each generation.

The aim of this paper is to introduce the modified version of the genetic algorithm. On the one hand, the proposed modification preserves the inherent advantages of genetic algorithms - the capability of exploring distant areas of feature space and avoiding getting stuck at local minima. On the other hand, it inherits the main advantage of the forward step-selection method, it is the possibility of setting the upper limit of the number of selecting features. The proposed approach is an alternative for the genetic algorithm proposed by Koller and Sahami [4] for the feature selection in the field of text recognition, but is much more straightforward and convenient to implement.

The paper presents the theoretical aspects of the proposed modification of the classic genetic algorithm and also results of its practical application for a set of real EEG data. The data set used in the research is a set submitted to the second BCI Competition (data set III - motor imaginary) by Department of Medical Informatics, Institute for Biomedical Engineering, Graz University of Technology [2]. The results of the feature selection obtained with the proposed algorithm were compared with the results obtained with the classic version of genetic algorithm. The methods were compared in terms of classification accuracy obtained with SVM classifiers using selected features.

2 Genetic Algorithms as a Method for Feature Selection

Genetic algorithm (GA) is one of the heuristic methods for solving the optimization problems. Nowadays, this method is used in many different research fields, but it originates from genetic sciences. To run the optimization process with a GA, first, the problem environment has to be defined, that is: a method of coding problem solutions to the form of GA individuals, a fitness function used for evaluating individuals in each generation, genetic operations used for mixing and modifying individuals, a method for selecting individuals and other additional GA parameters. The individuals of a GA are problem solutions and are composed of chromosomes and genes.

The most popular approach adopted when a GA is to be used in the feature selection process is a classic version of a GA, with one-point crossover and one-gene

mutation. The individual is composed of only one chromosome and its number of genes is equal to the number of features existing in the problem at hand. Each gene can take one of two binary values zero or one. Zero indicates that the feature of the number equal to gene index is not present in the solution coded in the individual, one - means that the feature is present in the current solution [1] [6].

In case of analyzing a data set of too low ratio of observations to features, launching the process of looking for the optimal subset of features from the full set of features or from the middle of this set (which is the case when a GA utilizes the uniform probability distribution while choosing the initial population), can cause that the removal of most of the features from the feature set can be impossible because of the direction of the fitness function. For example, this is the case when the selection process is guided directly by the classifier results. The problem is that the classifier optimization process is mostly aimed at maximizing the classifier precision. Removing some features from the feature set results in reduction of a number of classifier parameters, which mostly implies the reduction of the classifier precision. In this way, the fitness function based directly on a classifier precision prefers individuals containing more features. This is not desirable algorithm behavior because in general, the more parameters have to be trained, the lower generalization capabilities of the classifier.

Theoretically, the selection process has not to be guided purely by the classifier results. It is possible, for instance, to equip the GA fitness function with a penalty term, which will punish the individuals coding too many features. It is also possible to develop some specialized genetic operators converting such unwelcome individuals into individuals carrying a smaller number of features. In practice, however, the scale of the required reduction of the feature set is so large that it is extremely difficult to develop a stable function penalizing individuals carrying too many features or functions for converting these individuals. A much better solution seems to be to define such a way of coding individuals and such genetic operators which would allow searching the space of feature subsets with a limited number of features. The scheme of such an algorithm, proposed by the author of this paper, is given below.

1. An individual is composed of the number of genes equal to the maximal number of features. The maximal number of features is a GA parameter and is set by the user in relation to the number of observations, classifier mapping function and the overall problem characteristic.
2. Each gene can take an integer value from the interval $\{0, 1 \dots F\}$, where F denotes the dimension of the feature set. The initial value of the interval, value zero, was introduced on purpose - each time when it occurs in an individual, the number of classifier inputs is reduced by one.
3. Because of a very small ratio of number of genes to the total number of features, the basic genetic operation in proposed algorithm is mutation. Not only each individual in the population is mutated but also each gene of each individual. To be exact, an individual has a set of off-springs, each created by mutating one gene of the parent individual. Due to this, a large amount of new genetic information is introduced to population in succeeding algorithm generations.

4. The same factor and also a very aggressive version of mutation are responsible for changing the order of two basic GA steps - selection and reproduction. In the proposed approach, the reproduction takes place first. During this step the mother population is enlarged by adding new individuals created during the mutation and crossover. Next, the population is reduced to the initial size by selecting the individuals of the highest values of a fitness function. Due to the aggressive mutation a high variability of the population is achieved and due to reversing reproduction/selection steps best solutions are preserved (the selection is made not only between new created individuals but also between parent individuals, it is individuals which were the best in previous generation).
5. The fitness function is pure classifier accuracy.

The proposed algorithm permits individuals with duplicated features. Such solution was taken on purpose, because it allows evaluating individuals with a smaller number of features than the maximal number, set in the beginning of the algorithm. Each time when an individual with duplicated features appears (as a result of genetic operations or as a result of random initialization of a starting population), all its genes of the same value (apart from one) are set to zero. As a result, a smaller number of input variables is introduced to the classifier and it is possible to test whether higher classification precision cannot be obtained with a smaller number of features.

3 Experiments Settings

In order to verify the practical usability of the proposed genetic algorithm, a data set submitted to the second BCI Competition (data set III - motor imaginary) by Department of Medical Informatics, Institute for Biomedical Engineering, Graz University of Technology [2] was used. The data set was recorded from a normal subject (female, 25y) whose task was to control the movements of a feedback bar by means of imagery movements of the left and right hand. Cues informing about the direction in which the feedback bar should be moved were displayed on a screen in the form of the left and right arrows. The order of left and right cues was random. The experiment consisted of 280 trials, each trial lasted 9 seconds. The first 2s was quiet, at $t=2s$ an acoustic stimulus was generated and a cross + was displayed for 1s; then at $t=3s$, an arrow (left or right) was displayed as a cue. The EEG signals were measured over three bipolar EEG channels (C3, Cz and C4), sampled with 128Hz and preliminary filtered between 0.5 and 30Hz. The whole data set, containing data from 280 trials, was then divided into two equal subsets - the first one intended for classifier training and the second intended for external classifier test. Since only data from the first subset was published with target values (1 - left hand, 2- right hand), only this subset could be used in the process of classifiers training and testing.

The original data set, after removing the mean values from each canal, was transformed to a set of frequency band power features. The signal power was calculated separately for:

- 12 frequency bands: alfa band (8-13Hz) and five sub-bands of alfa band (8-9Hz; 9-10Hz; 10-11Hz; 11-12Hz; 12-13Hz); beta band (13-30Hz) and also five sub-bands of beta band (13-17Hz; 17-20Hz; 20-23Hz; 23-26Hz; 26-30Hz),
- each of 9 seconds of the trial,
- each of 3 canals (C3, Cz, C4).

Since the number of extracted features (equal to 324) was much too high in comparison to the number of observations (equal to 140), the feature selection process was performed. In order to reduce the number of features, two versions of genetic algorithm, described in Section 2, were used.

Because of a very adverse ratio of features to observations a classic linear SVM method was used in the classification process (in order to limit the overfitting phenomena). The classification threshold was set to 0.5 and hence all classifier results greater than 0.5 were classified as class "2" (right hand) and results smaller or equal to 0.5 were classified as class "1" (left hand). The classifiers accuracy was tested with 10-fold cross-validation. The final accuracy measure of a given feature set was the mean value calculated on the basis of classification accuracy obtained for all validation sets. The accuracy of one validation set was calculated according to the following equation:

$$A_k = \frac{R_k}{U_k}, \quad (1)$$

where: A_k - accuracy of k validation subset ($k=1\dots10$), R_k - number of properly classified cases from k validation subset, U_k - number of all cases in k validation subset.

4 Results and Discussion

4.1 Genetic Algorithm - Classic Approach

At the first stage of the survey the classic version of a genetic algorithm was used. The main parameters of the algorithm were set as follows: number of individuals: 50, number of chromosomes per individual: 1, number of genes per chromosome: 324; number of generations: 100. One individual coded one possible solution of the problem, it is one possible subset of the feature set. The initial population was chosen randomly with uniform probability distribution. Two classic genetic operations were used: one-point crossover (with the probability 0.8) and one-gene mutation (with the probability 0.025). Individuals proceeding to next population were selected according to the tournament method with the probability equal to 0.75. In order to avoid losing the best solution during the genetic operation or selection process, the record of the best individual was kept separately from the current population. After completing generation number 100, the algorithm was stopped and the feature subset coded in the individual of the highest value of the fitness function was accepted as the best one.

At first the algorithm was run only once and the feature subset presented in Table 1 was obtained. The mean classification accuracy of the initial population of the algorithm was equal to 68.06% and the accuracy of the best individual, encoding 153 features, was equal to 77.42%. After completing the algorithm, mean classification accuracy rised to 83.55% and the accuracy of the best individual, encoding 169 features, rised to 96.77%.

Table 1 A set of 169 features chosen by the classic genetic algorithm; value X indicates features which were encoded in the best individual (of accuracy equal to 96.77%), obtained in the first run of the algorithm

canal	C3									Cz									C4									
	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	
second	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	
8-13Hz	X			X	X	X		X				X			X	X			X	X	X		X		X		X	
8-9Hz	X	X			X		X	X	X						X				X	X	X	X					X	
9-10Hz		X	X	X						X	X	X	X	X		X	X									X	X	X
10-11Hz		X	X	X	X	X	X	X		X		X	X			X		X	X	X	X	X					X	
11-12Hz		X			X					X	X		X		X	X			X	X	X					X	X	
12-13Hz		X	X			X		X	X	X	X	X	X	X			X	X		X	X	X	X		X		X	
13-30Hz	X	X	X		X	X	X	X		X	X	X	X			X		X	X	X							X	X
13-17Hz	X	X	X		X					X	X		X	X			X	X	X		X	X	X	X	X		X	
17-20Hz	X				X	X	X	X		X	X	X			X	X					X						X	
20-23Hz	X		X	X	X		X	X	X		X			X			X			X	X	X	X	X	X			
23-26Hz					X	X	X	X				X	X	X		X		X	X	X	X	X		X				
26-30Hz	X					X	X	X								X	X		X	X		X			X		X	X

The result seemed to be rather odd not only because of the very high classification accuracy but especially because of the large number of features encoded in the best individual - even larger than the number of observations used in the survey (126 observation in each of the training sets). Such results have aroused serious doubts about the generalization capabilities of the classifier trained with the data set encoded in the chosen individual.

In order to find out whether choosing such a large set of features is a standard algorithm behavior, nine more runs of the algorithm were performed. Their results are presented in Table 2. As it can be observed, each run of the algorithm presents more or less similar behavior - the number of features chosen in succeeding runs (from 151 to 174) is always too large to obtain a classifier of high generalization capabilities.

Obviously, such algorithm behavior could be also a result of too small number of algorithm generations. In order to find out whether this is the case, the algorithm was run ten more times with enlarged training time (number of generation was set to 1000). After completing all ten runs, it occurred that there were no significant

changes in the size of feature subsets. Since at each run the same classification precision was obtained - 100%, further training of the algorithm was of no reason. Hence, two different conclusions of this experiment could be drawn - the first: all feature subsets chosen in each algorithm run are equally important, and the second: the algorithm, starting from a large number of features, has no tools for discarding redundant features and instead of this, adds to the set of really important features a lot of other features chosen randomly from the feature set. Which of these conclusions is the proper one, will be discussed in next section.

Table 2 Results obtained in succeeding algorithm runs (classic version); Mean [%] - mean accuracy obtained in the last generation, Best [%] - accuracy of the best individual, Number of features - number of features encoded in the best individual

No.	Mean [%]	Best [%]	Number of features
1	83.55	96.77	169
2	71.48	92.59	165
3	69.00	90.00	171
4	77.60	92.00	151
5	86.96	100.00	165
6	80.95	90.48	158
7	83.04	95.65	174
8	80.74	92.59	157
9	83.21	92.86	160
10	79.66	93.10	173

4.2 *Genetic Algorithm - Proposed Approach*

At the second stage of the survey the modified version of a GA, proposed in this paper, was run. In order to run the algorithm, the maximal size of feature subset had to be established. Taking into account the fact mentioned in Section 1, that in case of a linear classifier at least 10 training data per class per each input feature should be gathered in order to properly train the classifier [7], the maximal size of the feature set was set to 6 (each training set contained 126 observations). Other algorithm parameters were set as follows: number of individuals: 10, number of chromosomes per individual: 1, number of genes per chromosome: 6; number of generations: 50. The initial population was chosen randomly with uniform probability distribution. Two genetic operations were applied: one-point crossover (with the probability 1) and all-genes mutation (with the probability 1). Ten individuals of the highest value of a fitness function were chosen to next population. After completing generation number 50, the algorithm was stopped and the feature subset coded in the individual of the highest value of the fitness function was accepted as the best one.

This time, the algorithm was run at once 10 times. The results are presented in Table 3. Two facts can be noticed after taking a closer look at the Table 3. First, the classification accuracy obtained in succeeding algorithm runs was very high - it almost reached the same level as in the first experiment, when the feature subsets were composed of tens of features. So high accuracy, obtained with a classifier equipped with only 6 input variables, meant that the algorithm indeed chose the features significant for the classification process. Another very important fact was that the algorithm returned similar feature subsets. Only 15 of all 59 features encoded in all ten individuals did not repeat in different runs, remaining 44 genes encoded 11 features appearing in different combinations. Some of the features repeated even in more than 50% of algorithm runs (feature no. 104 in 8 runs, feature no. 24 in 7 runs and feature no. 5 in 6 runs). This meant that the algorithm behavior was stable and it also additionally underlined the fact that chosen feature subsets were of a high quality.

Table 3 Results obtained in succeeding algorithm runs (modified version); Mean 1st/last [%] - mean accuracy obtained in the first/last generation, Best 1st/last [%] - accuracy of the best individual of the first/last generation, F1, F2, F3, F4, F5, F6 - numbers of six features encoded in the best individual (the features descriptions can be found in Table 1, where first data cell denotes features no. 1, second data cell in the same row - feature no. 2 and so on)

No.	Mean 1st	Best 1st	Mean last	Best last	F1	F2	F3	F4	F5	F6
1	53.93	74.29	86.79	93.57	5	6	24	104	150	208
2	54.73	79.29	87.56	92.86	5	23	24	87	107	199
3	58.36	78.57	86.93	93.57	6	24	86	104	107	237
4	52.57	69.29	86.38	93.57	5	23	24	26	133	199
5	56.44	77.86	85.95	93.57	86	104	105	156	164	276
6	54.59	72.14	86.23	94.29	5	104	108	156	240	267
7	53.84	72.86	86.74	94.29	24	86	87	104	126	238
8	51.66	75	85.67	94.29		6	24	86	104	287
9	53.09	67.86	85.35	92.14	5	24	55	104	107	199
10	56.87	75.71	85.04	94.29	5	63	104	107	156	267

One could ask why the algorithm did not return exactly the same subset of features? The explanation is rather simple. Most of the extracted features were highly correlated (for example the features coding signal power in band 8-13Hz and features coding signal power in band 10-11Hz) and any swaps between them did not bring any significant difference in the classification precision. This means that all feature subsets, chosen in succeeding algorithm runs, should be regarded as equal in terms of feature significance. Of course if the BCI was to be created over the algorithm results, the set number 8 should be applied, since it is composed of the smallest number of features (5 features) and gives the highest rate of classification precision (94.29%).

Now it is the time to come back to the previous section, where the question was ask whether different feature subsets, chosen in different runs of the classic genetic algorithm, could be also regarded as subsets of the same importance. At this moment, the answer is straightforward: if six features are enough to obtain almost the same classification precision as 160 features, this means that about 150 features are redundant and have been chosen randomly (regardless of the optimization criterion). Hence, the subsets of features chosen in different runs of the classic GA cannot be regarded as equally important (most of all, because they can contain lots of unimportant features).

5 Conclusion

While comparing both algorithms, it should be underlined that the proposed modification of a classic genetic algorithm allows choosing a significantly smaller subset of important features without reducing the classification precision. The classification precision, exceeding 92% in all 10 cases, is really a good result. Moreover, this precision was obtained with linear SVM classifiers equipped with only 5 or 6 features, which allows to believe in high generalization capabilities of these classifiers.

Coming to future work, it should be stated that feature subsets, obtained with the proposed algorithm, were similar but not the same. It can mean that the features are highly correlated, as it was stated in the paper, but it can mean also that six features is still too much in the given problem and that the similar accuracy can be obtained with even a smaller number of features.

A second aspect which should be addressed now is how to force the algorithm to evaluate more feature subsets containing the smaller than maximal number of features? It seems that in order to deal with this task a specialized genetic operators aimed directly at reducing number of features should be developed and introduced to the algorithm.

References

1. Garrett, D., Peterson, D., Anderson, C., Thaut, M.: Comparison of Linear, Nonlinear, and Feature Selection Methods for EEG Signal Classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11(2), 141–145 (2003)
2. Data set III, II BCI Competition, motor imagery,
<http://bnci.de/competition/ii/index.html>
3. Jain, A., Duin, R., Mao, J.: A Review, Statistical Pattern Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1), 4–37 (2000)
4. Koller, D., Sahami, M.: Toward optimal feature selection. In: Proc. Machine Learning, pp. 284–292 (1996)
5. Lakany, H., Conway, B.: Understanding intention of movement from electroencephalograms. *Expert Systems* 24(5), 295–304 (2007)

6. Peterson, D., Knight, J., Kirby, M.: Anderson Ch, Thaut M Feature Selection and Blind Source Separation in an EEG-Based Brain-Computer Interface. EURASIP Journal on Applied Signal Processing 19, 3128–3140 (2005)
7. Raudys, S., Jain: A Small sample size effects in statistical pattern recognition: Recommendations for practitioners. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(3), 252–264 (1991)
8. Yom-Tov, E., Inbar, G.: Feature Selection for the Classification of Movements From Single Movement-Related Potentials. IEEE Transactions on Neural Systems and Rehabilitation Engineering 10(3), 170–177 (2002)

On the Use of Programmed Automata for a Verification of ECG Diagnoses*

Mariusz Flasiński, Piotr Flasiński, and Ewa Konduracka

Abstract. The recent results into constructing a formal model of a syntactic pattern recognition-based System for Teaching ElectroCardioGraphy (STECG) are presented. A class of programmed attributed regular grammars (PARG) is defined as a formal tool for a generation of ECG patterns. A programmed attributed finite-state automaton (PAFSA) is introduced for an analysis of ECG patterns. PAFSA is a basic formalism for a development of the STECG system.

1 Introduction

The idea of using computers for an ECG analysis was formulated in 1960s [24] and it was developed in 1970s [12]. Since then a lot of methods have been defined and a variety of computer systems for ECG diagnostics have been implemented. What is interesting, in an area of an ECG analysis both main approaches to pattern recognition are used. Apart from a "classical" decision-theoretic approach (including artificial neural networks) [1, 5, 6, 14–18, 21, 25], syntactic pattern recognition methods are widely applied [2, 13, 20, 23, 27–29]. The use of syntactic models results

Mariusz Flasiński

IT Systems Department, Jagiellonian University, ul. prof. St. Lojasiewicza 4, Cracow 30-348, Poland

e-mail: mariusz.flasinski@uj.edu.pl

Piotr Flasiński

Ernst & Young, Rondo ONZ 1, Warsaw 00-124, Poland

Ewa Konduracka

Institute of Cardiology, Collegium Medicum, Jagiellonian University, ul. Prądnicka 80, Cracow 31-202, Poland

* A preliminary phase of the research has been made by the author at the Department of Automatics and Biomedical Engineering, AGH University of Science and Technology, Al. Mickiewicza 30, Cracow 30-059, Poland.

from the fact that an ECG signal is treated in cardiology as a structure consisting of substructures such as e.g. a P wave, a QRS complex, a T wave, a PR segment, an ST segment etc. It means that syntactic pattern recognition, which treats a pattern as a complex structure, which can be decomposed into subpatterns that in turn are decomposed into simpler subpatterns, etc [4, 10, 11, 19] seems to be convenient especially for representing and analyzing ECG signals.

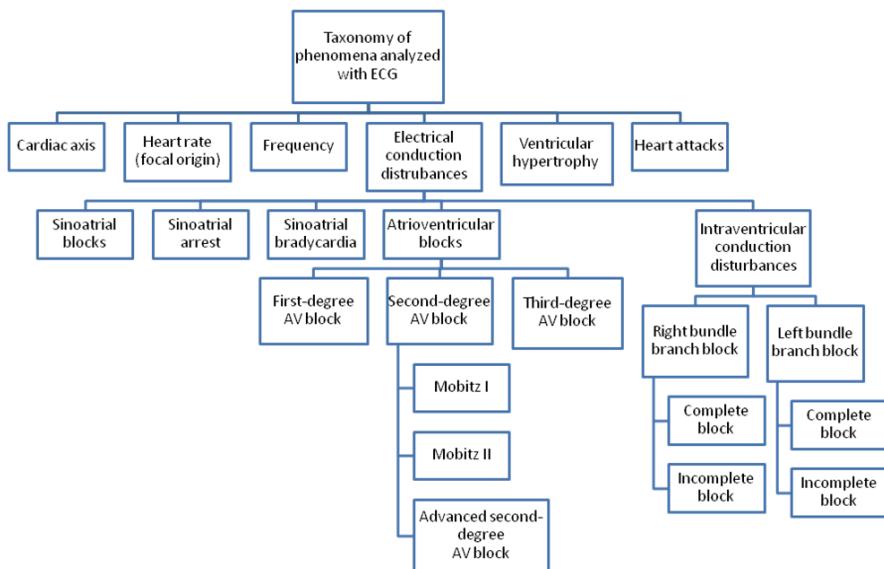


Fig. 1 A subclass-superclass - based ontology defined in the STECG project

A classification of an unknown object/phenomenon to one of a pre-defined classes/categories is the main goal of pattern recognition-based systems. In case of medical applications disease entities or abnormalities of organ functioning are defined as classes. Then, in both approaches the system should classify a representation of symptoms describing an unknown phenomenon to one of such classes. In case of a decision-theoretic approach phenomena are represented by feature vectors and they are assigned to one of clusters, which represents classes in a feature space by a classifier. In syntactic pattern recognition a phenomenon is represented as a symbolic structure belonging to a formal language and it is classified to a category by a formal automaton [4, 10, 11, 19]. If a classification task is concerned, none of the approaches prevails over another. However, if the system not only should classify an unknown phenomenon, but also characterize/describe its structural features, syntactic pattern recognition seems to prevail over "standard" pattern recognition approaches (probabilistic, discriminant function-based, neural networks, etc.).

The recent results of a research into constructing a formal model of the System for Teaching ElectroCardioGraphy, STECG conducted at IT Systems Department, Jagiellonian University, and Institute of Cardiology, Jagiellonian University

are presented in the paper. The main task of the system not only consists in verifying a diagnosis made by a student on a basis of an electrocardiogram, but also in assessing a diagnostic justification proposed by her/him. For the second purpose, the system should have an ability to characterize both structural and numerical features of an electrocardiogram. Therefore, a syntactic approach has been used for constructing the STECG system.

In section 2 we discuss a key issue of defining structural primitives, which are used for defining a syntactic representation of an electrocardiogram. A programmed attributed regular grammar, PARG is introduced in the third section. In section 4 we define a programmed attributed finite-state automaton, PAFSA, which is a basic formalism for implementing a syntax analyzer in the STECG system. The final section contains concluding remarks.

2 Definition of Structural Primitives

A functionality of the STECG system differs from typical pattern recognition-based ECG analysis systems. A classification of an unknown electrocardiogram to one of pre-defined classes is the main goal of such systems. STECG has to identify structural and numerical features of an electrocardiogram in order to verify a diagnostic justification proposed by a student. Therefore, the system has to simulate, somehow, a process of interpreting an electrocardiogram by a physician. In other words, STECG is a system designed on the basis of an image understanding paradigm [26] rather than the pattern recognition one. It means that for a design of STECG a formal conceptualization of the system domain, i.e. its ontology (in the Artificial Intelligence sense), should be defined. This ontology, which is based on a subclass-superclass (inheritance) relation, represents a taxonomy of phenomena considered in the STECG project. It determines a scope of the project as it is shown in Fig. 1.

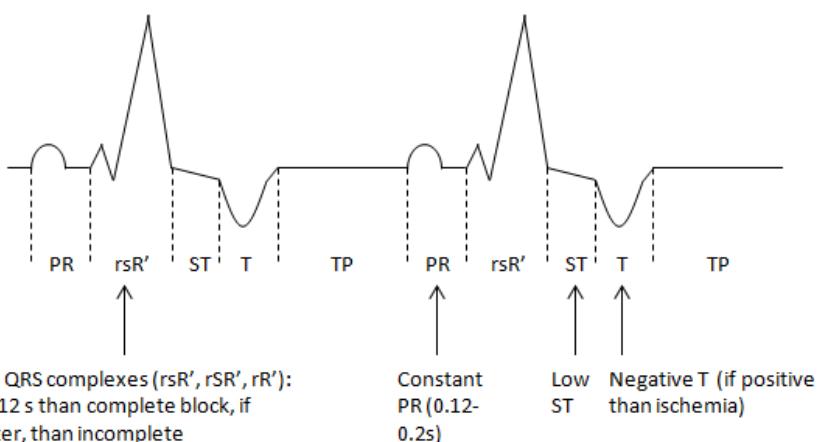


Fig. 2 A structural-parametric representation of a class *Right Bundle Branch Block*

Defining a set of structural *primitives*, which are adequate for a syntax analysis in an image understanding process has been the second key issue. As we have mentioned it in Introduction in syntactic pattern recognition complex patterns are decomposed into subpatterns that in turn are decomposed into simpler subpatterns, etc. The elementary patterns are called primitives and they are represented with symbols of a formal language alphabet. Such primitives have to be determined not only for a classification purpose (as it is made in typical ECG analysis systems), but also taking into account their adequacy for identifying structural ECG subpatterns that are essential in justifying a diagnosis. Therefore, a definition of primitives in the STECG project has been made on the basis of analyzing, together with a specialist in ECG diagnosing, a variety of electrocardiograms, which have been then described with the help of structural and parametric features. An example of such a description for a class *Right Bundle Branch Block*, *RBBB* (in V1 and V2 leads) is shown in Fig. 2. Some of primitives belong to a standard set of structural elements used for representing electrocardiograms. However, for a purpose of an image interpretation/understanding during a task of assessing a diagnostic justification certain additional (unique) primitives have to be defined. An example (sub)set of primitives identified for a class *Intraventricular Conduction Disturbances*, *ICD*, being a superclass of *Right Bundle Branch Block*, *RBBB* (cf. Fig. 1) is shown in Fig. 3.

As we can see in Fig. 2, a parametric description should be defined apart from identifying structural features for a purpose of ECG interpretation. Only then, a complete structural-parametric characterization of electrocardiograms, which allows the system to assess a diagnostic justification is possible. This parametric

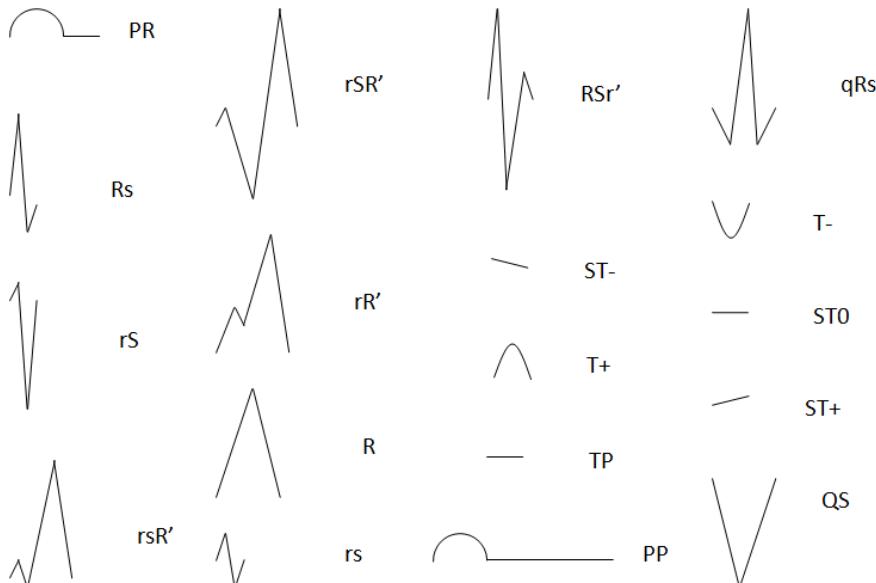


Fig. 3 A set of primitives for a class *Intraventricular Conduction Disturbances*

description is used for attributing a formal grammar and controlling its derivation process. These formalisms are introduced in the next section.

3 Programmed Attributed Regular Grammar

As we have discuss it in a previous section a structural representation of an electrocardiogram allows the system to asses a diagnostic justification proposed by a student. However, it is not sufficient to distinguish the difference between certain classes assumed in the ontology of the STECG project. For example, in a superclass *Right Bundle Branch Block*, *RBBB* to differentiate its subclass *Incomplete Block*, *IRBBB* from its subclass *Complete Block*, *CRBBB* (cf. Fig. 1) the system should check a parameter *QRS length*, which in the first case (*IRBBB*) belongs to an interval (0.10s, 0.12s), whereas in the second case (*CRBBB*) is greater than or equal to 0.12s. (The same holds for a superclass *Left Bundle Branch Block*, *LBBB*.)

Programming formal grammars allows one to enhance their generative power and, as a consequence, to enhance a discriminative power of corresponding automata. In standard programmed grammars, introduced by Rosenkrantz [22], such enhancement is achieved *via* controlling an order of production applications in a derivation process. Since, as we have discussed it above, distinguishing between certain classes of the STECG ontology is possible on the basis of numerical parameters only, programming based on primitive attributes should be applied for our grammar. It has been used for an image understanding with the help of graph grammars by Bunke [3]. We use his formalism of a production applicability predicate for a string regular grammar.

Let us define a programmed attributed regular grammar, PARG in the following way.

Definition 1. A programmed attributed regular grammar, PARG is a quadruple

$$G = (V, \Sigma, P, S)$$

where: V is a finite set of symbols, $\Sigma \subset V$ is a set of terminal symbols, $N = V \setminus \Sigma$ is a set of nonterminal symbols, P is a finite set of productions of the form:

$$(\pi : X \longrightarrow \alpha), \text{ in which}$$

$\pi : \mathcal{A} \longrightarrow \{\text{TRUE}, \text{FALSE}\}$ is the predicate of the production applicability, \mathcal{A} is a finite set of attributes, $X \in N$, $\alpha \in \Sigma \cup \Sigma N$, $S \in N$ is the starting symbol.

An application of a production in PARG is possible only, if its predicate π is fulfilled. Let us consider the following two productions for a class *Incomplete Right Bundle Branch Block*, *IRBBB*.

$$1. \pi = (l_{PR(i)} > 0.12) \wedge (l_{PR(i)} = l_{PR(i-1)}) : X^{(0)} \longrightarrow \mathbf{PR}X^{(1)},$$

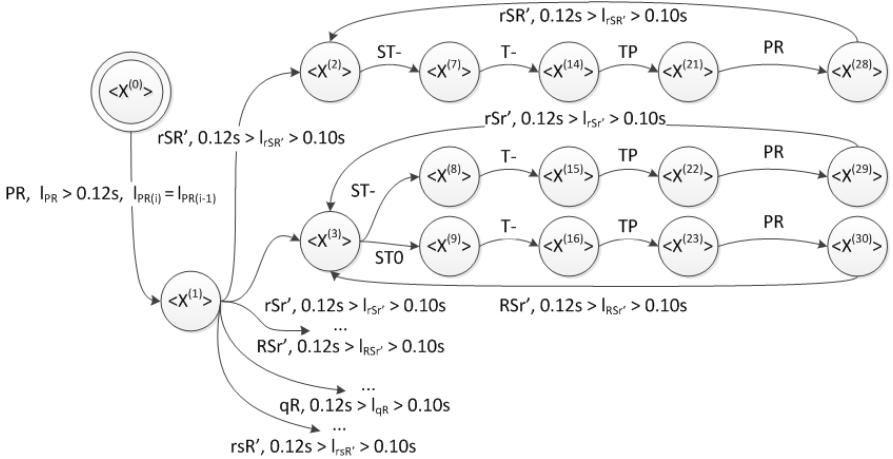


Fig. 4 A part of a programmed attributed finite-state automaton for a class *Incomplete Right Bundle Branch Block, IRBBB*

where $X^{(0)}, X^{(1)} \in N$, $\mathbf{PR} \in \Sigma$ is a terminal symbol representing a structural primitive corresponding to a PR segment (cf. Fig. 2). The predicate of the production applicability says that the length of the PR segment should be greater than 0.12s and its length should be (approximately) equal to the length of a previous PR segment (cf. Fig. 2).

$$2. \pi = l_{rSR'} \in (0.10s, 0.12s) : X^{(1)} \longrightarrow \mathbf{rSR}' X^{(2)},$$

where $X^{(1)}, X^{(2)} \in N$, $\mathbf{rSR}' \in \Sigma$ is a terminal symbol representing one of structural primitives (namely rSR') corresponding to a QRS segment (cf. Figs: 2 and 3). The predicate of the production applicability says that the length of the rSR' (QRS) segment should be greater than 0.10s and less than 0.12s (cf. Fig. 2).

Syntactic parts of PARG grammar productions have been defined with the help of a grammatical inference method of formal derivatives [10] in the STECG project.

4 Programmed Attributed Finite-State Automaton

Since we have defined a grammar allowing us to control a derivation process on a basis of attributes of structural primitives, we have applied a programmed attributed automaton as a formalism for implementing a syntax analyzer in the STECG system. A control table of the automaton can be defined easily, because there are simple rules of such a definition for regular grammars. Of course, a finite-state automaton has to be enhanced by a mechanism of a transition control. This enhancement has been made by a predicate of a transition permission.

Let us define a programmed attributed finite-state automaton, PAFSA in the following way.

Definition 2. A programmed attributed finite-state automaton, PAFSA is a quintuple

$$A = (Q, I, \delta, q_0, F)$$

where: Q is a finite nonempty set of states, I is a finite set of input symbols, δ is the transition function of the form:

$$\delta : Q \times I \times \Pi \longrightarrow Q, \text{ in which}$$

$\Pi : \mathcal{A} \longrightarrow \{\text{TRUE}, \text{FALSE}\}$ is the predicate of the transition permission, \mathcal{A} is a set of attributes, $q_0 \in Q$ is the initial state, $F \subseteq Q$ is a set of final states.

A definition of the predicate of the transition permission is straightforward. If a transition corresponds to a production of a PARG grammar, then a predicate of its permission is determined by the predicate of the applicability of this production. An example of a part of a programmed attributed finite-state automaton for a class *Incomplete Right Bundle Branch Block, IRBBB* is shown in Fig. 4. The first transition (from a state $X^{(0)}$ to a state $X^{(1)}$) corresponds to the first production introduced in a previous section. As it is easily seen both predicates of: the production applicability and the transition permission are equal. The same holds for the predicates of: the second transition (from a state $X^{(1)}$ to a state $X^{(2)}$) and the second production from a previous section.

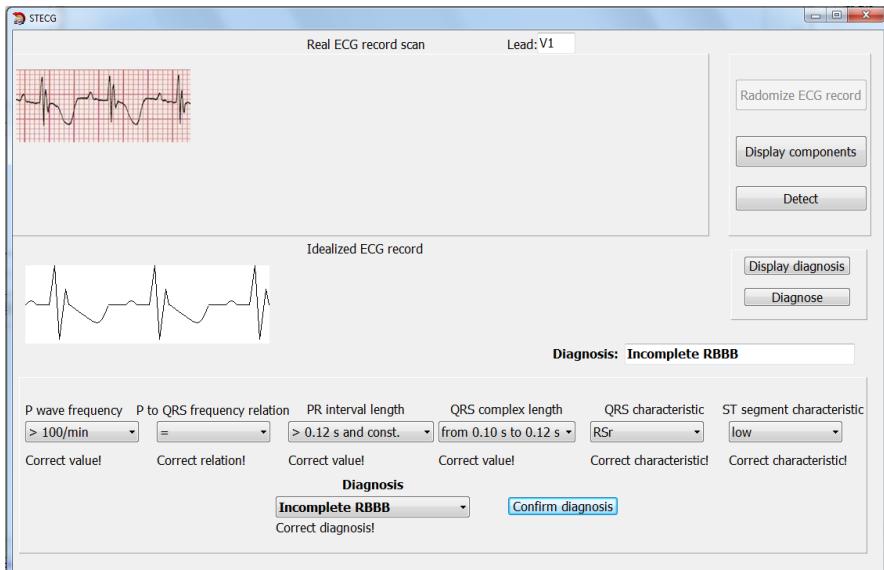


Fig. 5 An example of verifying an ECG diagnosis by the STECG system

5 Conclusions

The System for Teaching ElectroCardioGraphy, STECG² works in an interactive way. Firstly, an electrocardiogram is chosen randomly from the system library. Secondly, for an idealized ECG record a student proposes a diagnosis. Then, system suggests a set of structural and parametric characteristics and relations between them, which can be used for a diagnostic justification. The student has to choose adequate structural characteristics, proper values of parameters, correct relationships characteristics as a justification of the diagnosis. At the end, the system verifies them. An example of the system interface at the moment of verifying a diagnosis and a diagnostic justification is shown in Fig. 5.

The preliminary version of the system has been implemented for a part of the ontology shown in Fig. 1, which corresponds to a superclass *Electrical Conduction Disturbances, ECD*. Our experience concerning the use of syntactic pattern recognition in similar projects (e.g. in a project of constructing a system for an analysis of an auditory brainstem response, ABR [8]) shows that in case of extending the scope of an application domain of the system an enhancement of a formal model is usually necessary. Fortunately, such an enhanced model based on context-free dynamically programmed grammars and automata, a DPLL(k) model, has been constructed recently at IT Systems Department, Jagiellonian University in Cracow [7,9]. The next results of the research into the System for Teaching ElectroCardioGraphy, STECG will be a subject of further publications.

Acknowledgements. The authors thank Professor Piotr Augustyniak for valuable discussions during a preliminary phase of the project.

References

1. Barro, S., Fernandez-Delgado, M., Villa-Sobrino, J.A., Regueiro, C.V., Sanchez, E.: Classifying multichannel ECG patterns with an adaptive neural network. *IEEE Eng. Med. Biol. Mag.* 17, 45–55 (1998)
2. Belferte, G., De Mori, R., Ferraris, F.: A contribution to the automatic processing of electrocardiograms using syntactic methods. *IEEE Trans. Biomed. Eng.* 26, 125–136 (1979)
3. Bunke, H.O.: Graph grammars as a generative tool in image understanding. In: Ehrig, H., Nagl, M., Rozenberg, G. (eds.) *Graph Grammars 1982. LNCS*, vol. 153, pp. 8–19. Springer, Heidelberg (1983)
4. Bunke, H.O., Sanfeliu, A. (eds.): *Syntactic and Structural Pattern Recognition - Theory and Applications*. World Scientific, Singapore (1990)
5. Dong, J., Xu, S., Zhan, C.: ECG recognition and classification: approaches, problems and new method. *J. Biomed. Eng.* 24, 1224–1229 (2007)
6. Engin, M.: ECG beat classification using neuro-fuzzy network. *Patt. Rec. Lett.* 25, 1715–1722 (2004)

² A preliminary version of the system has been implemented by Piotr Flasiński at the Department of Automatics and Biomedical Engineering, AGH University of Science and Technology in Cracow.

7. Flasiński, M., Jurek, J.: Dynamically programmed automata for quasi context sensitive languages as a tool for inference support in pattern recognition-based real-time control expert systems. *Pattern Recognition* 32, 671–690 (1999)
8. Flasiński, M., Reroń, E.z., Jurek, J., Wójtowicz, P., Atlasiewicz, K.: Mathematical linguistics model for medical diagnostics of organ of hearing in neonates. In: Wyrzykowski, R., Dongarra, J., Paprzycki, M., Waśniewski, J. (eds.) *PPAM 2004. LNCS (LNAI)*, vol. 3019, pp. 746–753. Springer, Heidelberg (2004)
9. Flasiński, M., Jurek, J.: On the analysis of fuzzy string patterns with the help of extended and stochastic GDPLL(k) grammars. *Fundamenta Informaticae* 71, 1–14 (2006)
10. Fu, K.S.: *Syntactic Pattern Recognition and Applications*. Prentice Hall, Englewood Cliffs (1982)
11. Gonzales, R.C., Thomason, M.G.: *Syntactic Pattern Recognition: An Introduction*. Addison-Wesley, Reading (1978)
12. Horowitz, S.L.: A syntactic algorithm for peak detection in waveforms with applications to cardiography. *Comm. ACM* 18, 281–285 (1975)
13. Koski, A., Juhola, M., Meriste, M.: Syntactic recognition of ECG signals by attributed finite automata. *Pattern Recognition* 28, 1927–1940 (1995)
14. Maglaveras, N., Stamkopoulos, T., Diamantaras, K., Pappas, C., Strintzis, M.: ECG pattern recognition and classification using non-linear transforms and neural networks: a review. *Int. J. Med. Inform.* 52, 191–208 (1998)
15. Martis, R.J., Chakraborty, C., Ray, A.K.: A two-stage mechanism for registration and classification of ECG using Gaussian mixture model. *Pattern Recognition* 42, 2979–2988 (2009)
16. Noponen, K., Kortelainen, J., Seppanen, T.: Invariant trajectory classification of dynamical systems with a case study on ECG. *Pattern Recognition* 42, 1832–1844 (2009)
17. Osowski, S., Linh, T.H.: ECG beat recognition using fuzzy hybrid neural network. *IEEE Trans. Biomed. Eng.* 48, 1265–1271 (2001)
18. Pan, J., Tompkins, W.J.: A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* 32, 230–236 (1985)
19. Pavlidis, T.: *Structural Pattern Recognition*. Springer, New York (1977)
20. Papakonstantinou, G., Skordalakis, E., Gritzali, F.: An attribute grammar for QRS detection. *Pattern Recognition* 19, 297–303 (1986)
21. Piętka, E.: Feature extraction in computerized approach to the ECG analysis. *Pattern Recognition* 24, 139–146 (1991)
22. Rosenkrantz, D.J.: Programmed grammars and classes of formal languages. *Journal of the Association for Computing Machinery* 16, 107–131 (1969)
23. Skordalakis, E.: Syntactic ECG processing: a review. *Pattern Recognition* 19, 305–313 (1986)
24. Stallmann, F.W., Pipberger, H.V.: Automatic recognition of electrocardiographic waves by digital computer. *Circ. Res.* 9, 1138–1143 (1961)
25. Sternickel, K.: Automatic pattern recognition in ECG time series. *Comp. Meth. Programs in Biomedicine* 68, 109–115 (2002)
26. Tadeusiewicz, R., Ogiela, M.R.: *Medical Image Understanding Technology*. Springer, Heidelberg (2004)
27. Trahanias, P., Skordalakis, E.: Syntactic pattern recognition of the ECG. *IEEE Trans. Patt. Analysis Mach. Intell.* 12, 648–657 (1990)
28. Tumer, M.B., Belfore, L.A., Ropella, K.M.: A syntactic methodology for automatic diagnosis by analysis of continuous time measurements using hierarchical signal representations. *IEEE Trans. on Syst. Man Cybern.* 33, 951–965 (2003)
29. Udupa, J., Murthy, I.S.N.: Syntactic approach to ECG rhythm analysis. *IEEE Trans. Biomed. Eng.* 27, 370–375 (1980)

Blood Flow Modeling in a Synthetic Cylindrical Vessel for Validating Methods of Vessel Segmentation in MRA Images

Grzegorz Dwojakowski, Artur Klepaczko, and Andrzej Materka

Abstract. The paper presents new concept of validating blood segmentation methods. The idea is to create a Magnetic Resonance Angiography simulator which can generate synthetic MRA images based on pattern objects. In order to reproduce angiographic sequences such as Time of Flight or Susceptibility Weighted Imaging, laminar flow in digital phantoms was modeled. Results were presented and discussed. Comparison of the new concept with existing validation methods was carried out.

1 Introduction

Magnetic Resonance Imaging (MRI) is one of the most popular radiology techniques used to obtain anatomical images of the human body [4]. Modern MRI scanners can receive high resolution 3D images with good contrast between different tissues. In comparison to traditional X-rays or Computer Tomography, MRI is non-invasive because it does not use ionizing radiation. Another advantage is the number of different sequences. Using angiographic techniques such as Time of Flight [3] and Susceptibility Weighted Imaging [18] methods combined together, results in a full map of veins and arteries [9]. Such a map carries important information about patient health and can be used in diagnosis and planning surgical operations.

Image segmentation methods are created in order to help physicians in analyzing those complex images [6]. Separating vessels from other tissues can result in revealing important information. Reconstructing vessel walls may lead to detecting clots, narrowings or other anomalies much faster than without image processing

Grzegorz Dwojakowski · Artur Klepaczko · Andrzej Materka
Lodz University of Technology, Institute of Electronics,
ul. Wolczanska 211/215, 90-924 Lodz
e-mail: grzegorz.dwojakowski@gmail.com,
{aklepaczko,materka}@p.lodz.pl

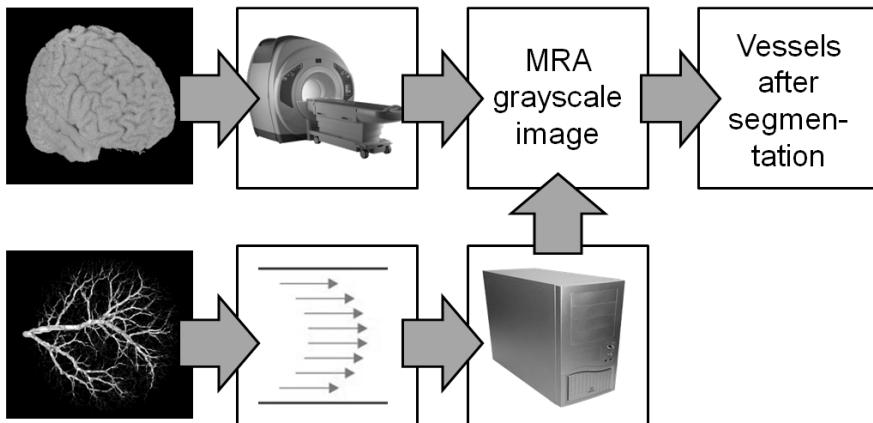


Fig. 1 Comparison of standard MRI imaging sequence with described concept

techniques [21]. Although results of blood segmentation algorithms can increase the scope of knowledge about patient health, they are rarely used in clinics. Diagnosticians cannot rely on segmented images and must make diagnoses based on their knowledge and medical experience. This is mainly due to problems with validating results of blood segmentation methods [13]. Their results cannot be compared with some ground truth pattern. Firstly, after transformation and rasterization of input data during the imaging sequence, a significant part of information is lost irreversibly. Secondly, there is no method to get precise dimensions of a patient vessel network. It is not possible to measure such complex structures inside a human body.

In order to use image processing methods for medical purposes there is a need of creating a reliable validation method. A method which not only allows to evaluate the correctness of reconstructed structures but also helps in the process of developing segmentation algorithms. To meet these requirements we propose a new method of validation. We want to simulate angiographic imaging sequences using a computer unit (cf. Fig. 1). This allows us to replace real image data with a 3D digital model of human vessels. A properly implemented simulator working on known patterns will create an effective and objective criterion of validation for image segmentation methods.

In standard MRI imaging, contrast is acquired thanks to chemical and physical differences between tissues. However, in magnetic resonance angiography (MRA) sequences voxel brightness is also dependent on the position of proton particles in time. Here, the phenomenon of blood flow is utilized. During the imaging process blood is moving in vessels while other tissues are stationary. Therefore, an appropriate simulation of this phenomenon is crucial for the postulated approach to succeed.

At an early stage of this project development laminar flow was modeled and applied in a simple digital tubular phantom vessel. The reason for this is the existence of real test objects with exactly the same parameters. After performing blood flow simulation for digital phantom there is a possibility to measure values obtained in a physical model and to compare the results. This phantom can be used not only to test the correctness of blood flow simulation but also to validate results of MRA sequences. If successful this could lead to using more complex vessel models, similar to the ones in a human body.

2 Blood Vessel Segmentation

Image segmentation techniques are used to divide data into regions with common characteristics. In case of MRA images the purpose is to separate blood vessels from other tissues. As mentioned earlier, it allows to portray arteries and veins as a 3D model. Data in that form is much easier to be analyzed than 2D cross-sections. Additionally, segmentation allows to automatically search for risk regions based on vessel diameter.

There are plenty of segmentation algorithms to choose from [15]. Depending on which part of the body has been imaged and what types of vessels need to be detected different methods should be applied. We can divide vessel segmentation techniques into two main categories: the mathematical morphology approach and geometrical methods [10]. Algorithms from the first group are considered to be fast but not very effective. The best example is image thresholding or a more advanced technique known as region growing. Output images are obtained almost instantly, but—e.g in cerebrovascular imaging—it is impossible to separate small veins from white matter or even from the bone. There are more advanced morphology methods such as watershed segmentation or top-hat.

Algorithms from the second group not only use the voxel brightness level but also the relative position of neighboring points. Deformable models or generalized cylinders approach are thought to give accurate results. However, the time needed to process a high resolution MRA 3D image is very long and the reconstruction of veins with small diameters (compared to the size of the voxel) is impossible. Base on experiments, low level geometric methods seem to be more appropriate. Usage of Hessian matrix to find local direction of white tubular objects is one of these methods [11].

In order to test various attempts and develop new ones, an image segmentation application was created. Using implemented image functions and filters there is a possibility to build and run complex algorithms in a short time. Additionally, thanks to an intuitive graphic user interface (see Fig. 2) and advanced task tracking system, changing used parameters or adding new methods to existing process posses no problem.

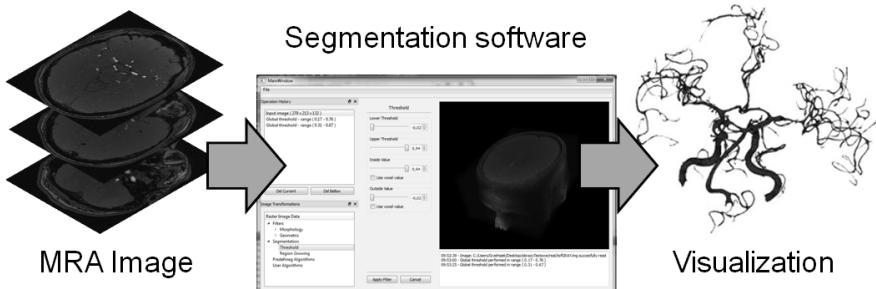


Fig. 2 Graphical user interface of created segmentation software

3 Validation of Biomedical Image Processing Methods

Although some segmentation algorithms results looks promising it is difficult to perform an objective assessment of their correctness [8]. The Most common method of validation is relying on medical knowledge and subjective assessment made by physicians [5]. However, gathered information cannot be used as a reliable pattern due to uncertainty and poor reproducibility.

The Second group of validation methods is based on physical phantoms. These artificial structures are mainly used to calibrate MRI scanners, but can also be used to test results of segmentation methods. This time, the pattern is familiar and comparison can be made. This method of validation is much more accurate, but has its drawbacks. Physical phantoms are expensive and, so far, there is no structure similar to the real blood network. This is a result of complicated topology and small vessel diameters.

The third group consists of digital phantoms. Similarly to physical structures they provide a priori known and measurable reference pattern [1]. A complex geometrical shape is much easier to create using computer 3D graphics. What is more, they are much faster and cheaper to create. They can be duplicated with different parameters and can be used to create large sets of test objects. The only disadvantage is that it is not possible to use these phantoms in a real MRI scanner.

All three methods have their drawbacks and potentials. An ideal solution seems to be a combination of all methods. As a pattern we would like to use easy to create digital objects, with known dimensions and blood flow inside it.

4 MRA Simulation

Problems with validating segmentation results can be solved by an MRA simulator. Simulating movement of blood in vessels allows implementing angiographic sequences and obtaining images similar to those from a real scanner (with noise and

distortion caused by imaging sequence). Contrary to image acquisition using real MRI apparatus, the input object is known and can be used as a reference source.

Simulating the physical phenomena of an MRI scanner is a complex problem. Only a few attempts to this issue can be found in literature [2, 14, 16, 17], but they work only for static objects. In case of angiographic imaging, these methods cannot be used because sequences such as Time of Flight (ToF) and Susceptibility Weighted Imaging (SWI) are based on blood flow in arteries and veins. In ToF, image contrast is acquired by unsaturated molecules of blood which flows through acquisition volume in a given time moment. SWI depends on blood oxygenation and uses amplitude and phase gradient echo, with compensated blood flow effect.

5 Blood Flow Modeling

Simulation of laminar flow is performed in COMSOL Multiphysics software by solving incompressible Navier-Stokes equations. This type of blood movement, known also as stratified flow, assumes that fluid moves in parallel layers [12]. Each layer has its own speed and slides past one another so there is no lateral mixing. Blood flows in one direction. Velocity is greatest in the middle of the cylinder. The value decreases as we approach the vessel wall. For simulating blood flow in a pipe this model is sufficient. For more complex structures with many branches and bifurcation there is a need to use turbulent model of flow.

To solve flow equations there is a need to set viscosity and density of the liquid. Based on literature these values for blood were adjusted to 1060 Kg/m³ (density) and 0.005 Ns/m² (dynamic viscosity).

The goal of this part of the project is to simulate blood flow in a complex artificial model of a vessel tree. It cannot be done until it is certain that the simulation is consistent with the real flow. That is the reason why the geometric shape of a digital pattern should be uncomplicated and identical to the physical phantom. If successful this could lead to using vessel models, similar to these in a human body.

Firstly, Geometry objects are created on the model of the Flow Phantom Set produced by Shelley Medical Imaging Technologies [20]. This Model is compatible with the high class CompuFlow 1000 MR pump which allows to force and precisely control the flow in test objects [19].

This Phantom Set consists of 4 straight (Fig. 3a) and 1 U-bend tubes (Fig. 3b). The geometrical parametrs of the tubes are as follows:

- Straight tubes phantom QA-STV (all tubes are 207 mm long)
 - 8 mm diameter vessel with 75% sinusoidal stenosis by diameter
 - 8 mm diameter vessel
 - 8 mm diameter vessel with 50% sinusoidal stenosis by diameter
 - 5 mm diameter vessel
- U-bend phantom QA-USV
 - 8 mm diameter, 44 mm midline radius-of-curvature

There is no fluid leak through the walls and no deformation caused by flow. This simplification makes the model easier to implement. Created digital models are an exact copy of their real equivalents, including details such valves and silicone control drillings (cf. Fig. 4).

6 Results

Blood flow simulation was performed for both phantom sets. Fluid flow is forced by setting pressure difference between input (11208 Pa) and output (11148 Pa) boundaries [7]. There is no slip allowed through phantom walls. Distribution of pressure and velocity of particles inside straight vessels are depicted in Fig. 5 and 6. In a vessel with a 5 mm diameter pressure distribution is linear. According to the theory of laminar flow, highest velocity is obtained in the center of the tube. Moving towards the wall this value decreases. Based on the laminar flow solution, particle tracing was performed. The Number of molecules was set to 32. All of them were defined by density and diameter. At the beginning, all particles are located at input boundary. With time, each element moves towards the exit. All trajectories are straight lines parallel to the main axis of the vessel. The fastest particles reach its

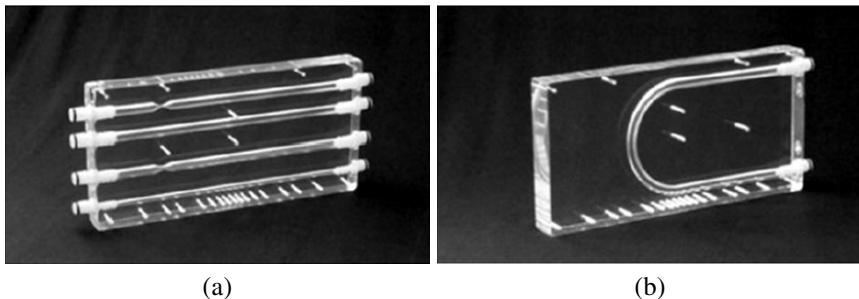


Fig. 3 Physical silicon phantoms: a) straight cylinders with stenosis, b) U-bend tube

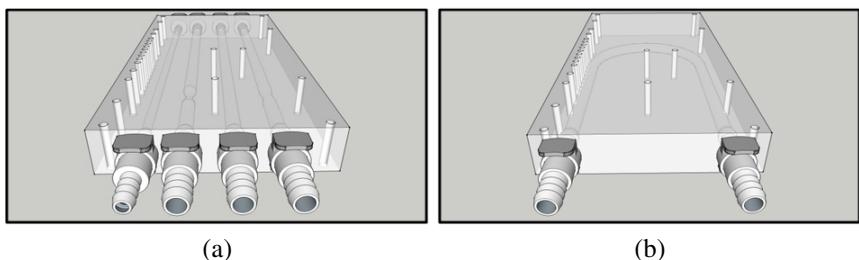


Fig. 4 Digital phantoms based on respective real models illustrated in Fig. 3

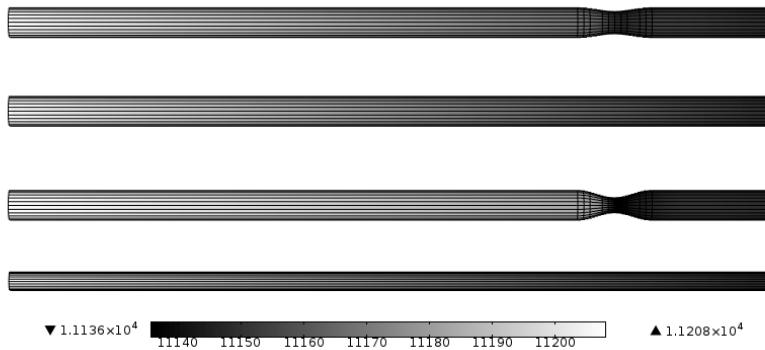


Fig. 5 Pressure distribution in 4 tubular phantoms. Value range is 11136–11208 Pa.

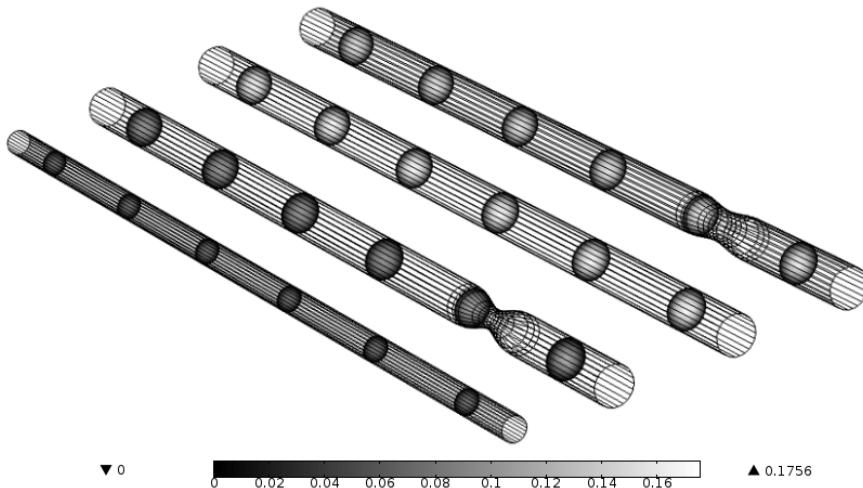


Fig. 6 Velocity magnitude in 4 tubular phantoms. Value range is 0–0.18 m/s.

goal in about 2.5 second; ones next to the wall are over five times slower. Results from this study are relevant to values obtained in laminar flow simulation.

Similar situation can be observed for a second straight tube of 8 mm. The main difference is time of flight of particles in the middle of the tube which equals 1.19 seconds. These values are relevant to the theory where local velocity depends on distance from the edge of the vessel.

Next two phantoms contain sinusoidal stenosis. Parameters of the flow are still the same. Due to narrowing, pressure distribution is no longer a linear function. Although in a phantom with 75% stenosis values of magnitude velocity are similar to those in a straight tube, small acceleration is visible in the narrowing region. The fastest particle reaches a distance of 207 mm in 1.25 s which is 0.06 slower

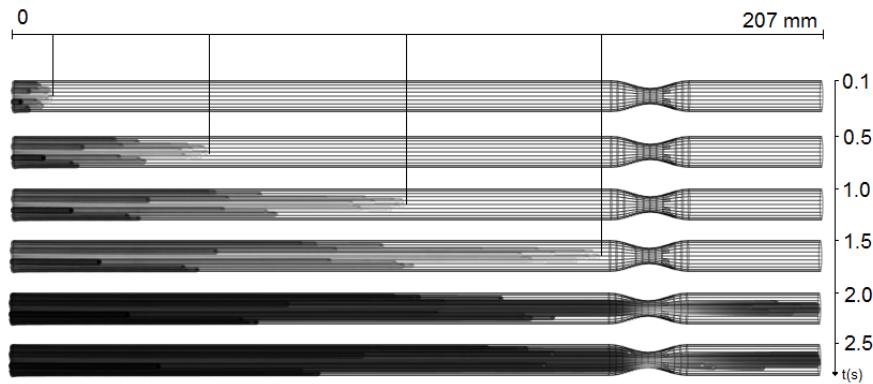


Fig. 7 Particle trajectories in vessel with 50% stenosis by diameter

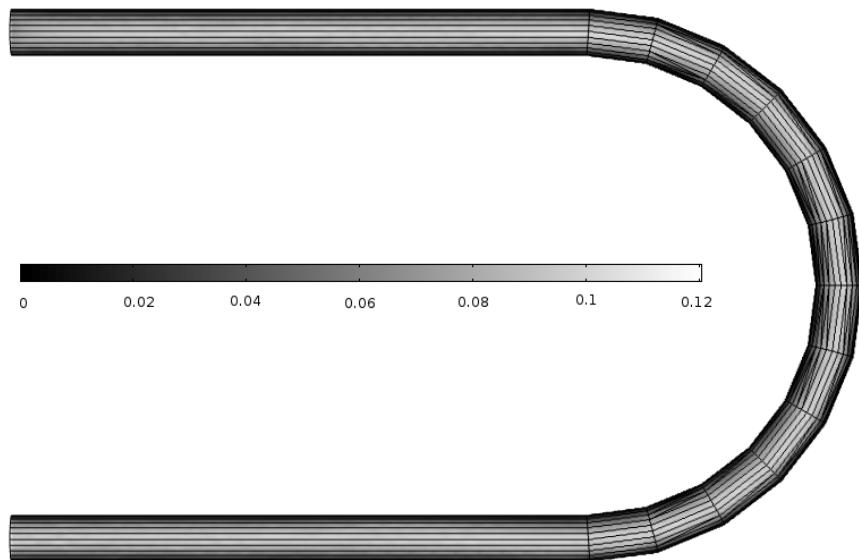


Fig. 8 Velocity magnitude in U-bend phantom. Value range is 0–0.12.

than in previous test. In the case of a vessel with 50% narrowing (Fig. 7), pressure applied on the particles before reaching stenosis is almost constant and equals the input parameter. Velocity magnitude is no longer a constant value. It changes from 0.2 m/s in the input section to about 0.4 m/s in the middle of stenosis. Time of the flow is much longer and equals 1.79 s. After the narrowing, the density of particles is higher because they do not return to their previous positions. The final U-bend phantom is different from previous 4 tubes. It is much longer and bent to an angle of 180 degrees. Because of the lack of narrowings, pressure is a linear function.

Time of flight for the particles in the middle equals 3.55 s. In the curve section small deviations from the central axis can be observed in the molecules trajectory (cf. Fig. 8).

7 Conclusion

A new concept of validating blood segmentation methods was presented. The idea was compared with existing methods. For the purpose of implementing angiographic sequences blood, flow was simulated in digital phantoms specially created for this project. Expected values and 3D coordinates were gathered and will be used in future validation process. Although much work still needs to be done, this report brings us closer to the final goal of a project—an MRA Simulator.

Acknowledgements. This paper was supported by the Polish National Science Centre grant no.

N N519 650940.

References

1. Aubert-Broche, B., Griffin, M., Pike, G., Evans, A., Collins, D.L.: IEEE Transactions on Medical Images 25, 1410–1416 (2006)
2. Benoit-Cattin, H., Collewet, G., Belaroussi, B., Saint-Jalmes, H., Odet, C.: Journal of Magnetic Resonance 173, 97–115 (2005)
3. Bernstein, M.A., Huston, J., Lin, C., Gibbs, G.F., Felmlee, J.P.: Magn. Reson. Med. 46(5), 955–962 (2001)
4. Bogorodzki, P.: Przeglad Elektrotechniczny 9, 40–45 (2009) (in Polish)
5. Castro, F., Pollo, C., Meuli, R., Maeder, P., Cuisenaire, O., Cuadra, M., Villemure, J., Thiran, J.: IEEE Transactions on Medical Images 25, 1440–1448 (2006)
6. Chmielewski, L., Kulikowski, J., Nowakowski, A.: Obrazowanie biomedyczne, vol. 8. Akademicka oficyna wydawnicza EXIT (2003) (in Polish)
7. Cieslicki, K.: Hydrodynamiczne uwarunkowania krazenia mozgowego. Akademicka Oficyna Wydawnicza EXIT (2001) (in Polish)
8. Crum, W., Camara, O., Derk, L., Hill, G.: IEEE Transactions on Medical Images 25, 1451–1461 (2006)
9. Deistung, A., Dittrich, E., Sedlacik, J., Rauscher, A., Reichenbach, J.R.: J. Magn. Reson Imaging 29(6), 1478–1484 (2009)
10. Eiho, S., Qian, Y.: Proc. Computers in Cardiology 7, 525–528 (1997)
11. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
12. Jezowiecka-Kabsch, K., Szewczyk, H.: Mechanika plynów. Oficyna Wydawnicza Politechniki Wrocławskiej (2001) (in Polish)
13. Jonnin, P., Krupinski, E., Warfield, S.: IEEE Transactions on Medical Images 25, 1405–1409 (2006)

14. Jurczuk, K., Kretowski, M.: *Biocybernetics and Biomedical Engineering* 29, 31–46 (2009)
15. Kirbas, C., Quek, F.: *ACM Comput. Surv.* 36(2), 81–121 (2004)
16. Kwan, R.S., Evans, A., Pike, G.: *IEEE Transactions on Medical Imaging* 18, 1085–1097 (1999)
17. Petersson, J., Christoffersson, J., Golman, K.: *Magnetic Resonance Imaging* 11, 557–568 (1993)
18. Reichenbach, J.R., Haacke, E.M.: *NMR in Biomedicine* 14, 453–467 (2001)
19. Shelley, M.I.T.: *Cardioflow 1000* MR computer controlled flow system (2013)
20. Shelley, M.I.T.: *MRI quality assurance flow phantom set* (2013)
21. Strzelecki, M., Materka, A., Kocinski, M., Szczyplinski, P., Deistung, A., Reichenbach, J.: *Acta Bio-Optica et Informatica Medica. Inzynieria Biomedyczna* 16, 167–172 (2010) (in Polish)

Swarm Optimization and Multi-level Thresholding of Cytological Images for Breast Cancer Diagnosis

Marek Kowal, Paweł Filipczuk, Andrzej Marciniak, and Andrzej Obuchowicz

Abstract. This paper presents a novel approach for multi-level thresholding of cytologic images. Typically, thresholding is applied in order to segment the image into regions of interest or objects, each having a high level of homogeneity in some parameter such as luminance. Homogeneous regions are then used to generate a set of features discriminating categories occurring in a given diagnostic problem. Instead of homogeneity measure, our approach uses a classifier to evaluate the quality of segmentation solution directly. The candidate solutions (sets of threshold values) are generated with use of the stochastic swarm intelligence-based metaheuristics. Experimental results demonstrate the promising performance of the proposed classification-driven segmentation in application to breast cancer diagnostics.

1 Introduction

Recently there has been an increase in interest in computer-aided cytology. Several researchers have studied the segmentation of cytological images of breast tumors, proposed new features or classification algorithms [6, 7, 10, 13–16, 21].

Most of the contemporary computer-aided cancer diagnosis systems are based on the assumption that a nucleus of the cell is the place where a breast cancer malignancy can be observed. Hence, it is crucial for any camera-based automatic diagnosis system to separate nuclei from the rest of the image. Correct extraction of nuclei can be very difficult due to noise, nuclei overlapping or high variation of image parameters (outcoming from changing lighting conditions, tissue staining etc.). In this paper, a new approach not requiring detection of any real objects (such as nuclei or cytoplasm) has been proposed.

Marek Kowal · Paweł Filipczuk · Andrzej Marciniak · Andrzej Obuchowicz
Institute of Control & Computation Engineering, University of Zielona Góra, Ogrodowa 3b,
65-246 Zielona Góra, Poland
e-mail: {M.Kowal,P.Filipczuk,A.Marciniak,
A.Obuchowicz}@issi.uz.zgora.pl

We assume that malignancy can be discriminated from the morphometric features describing homogeneous regions in the image, apart from the fact whether these regions are corresponding to any semantic objects or not. There might be an implicit mapping between regions determined by clusters in luminance space and classes of cancer, good enough for practical implementation in diagnostic decision supporting system. We have made an attempt to find this mapping using three swarm optimization algorithms, where the adjustment of thresholding parameters was driven by the feedback from classifier.

The paper is divided into 4 sections. Section 1 presents an introduction into breast cancer diagnosis. Section 2 describes the proposed approach and provides basic concepts about swarm optimization algorithms used in this work. Experimental investigations are shown in Section 3. The paper ends with our conclusions.

2 Multi-level Thresholding and Swarm Optimization

Image thresholding is the key technique in image processing field [8, 18, 19]. Frequently, the criterion used to select the optimal set of thresholds is based on the image histogram [11, 17]. However, histogram-based criteria can not guarantee obtaining high diagnostic accuracy. Instead of homogeneity measure a predictive accuracy is used to evaluate threshold sets. Objects extracted in this manner usually have no medical interpretation but hypothesis arises that their morphometric features can discriminate benign from malignant cases.

Proposed scheme of threshold evaluation is similar to wrapper strategy usually applied to feature selection [12]. In the first step, threshold set to be evaluated is applied to segment 225 benign and 225 malignant images. Next, segmented objects are measured using 3 features: ratio of the area of the region to the total area of the image, ratio of the perimeter of the region to the sum of perimeters of all regions, ratio of the Euler number of the region to the sum of Euler numbers of all regions. For each image $n_f = 3(K + 1)$ features are calculated, where K is the number of thresholds. Finally images are classified using k-nearest neighbors (kNN) classifier with parameter $k = 7$ chosen experimentally [2]. Predictive accuracy of kNN classifier was estimated using the n-fold cross-validation technique [3]. The fold was a set of 9 images representing 1 patient. This means the images belonging to the same patient were never at the same time in the training and testing set.

Searching for optimal thresholds is NP-hard problem and classical methods are applicable only when the set of thresholds is small enough. Moreover, since the wrapper strategy must train classifier 50 times (because cross validation is used) for each threshold set, the evaluation of objective function becomes computationally very expensive. To overcome this problem swarm intelligence (SI) algorithms as searching strategies are applied to find optimal thresholds [1, 4, 20, 22]. Three promising algorithms, i.e., ant colony optimization (ACO), honey bee mating optimization (HBMO) and firefly algorithm (FA) were chosen to solve the problem and are described in detail further in this section.

2.1 Ant Colony Optimization

In the early 1990s, ACO was introduced by Marco Dorigo [4] as a novel nature-inspired meta-heuristic, which underlying idea was to use several constructive computational agents simulating the behavior of real ants. While an isolated ant moves practically at random (exploration force), an ant encountering a previously laid pheromone trail can detect it and decide with high probability to follow it and consequently reinforce the trail with its own pheromone (exploitation force).

After initialization of pheromone values, the main loop consists of three iterative main steps. In the following, we give a more detailed description of ACO applied to the problem of multi-level image thresholding.

Initialize. At the start of algorithm, parameters are set, i.e. all pheromone values are initialized to a value τ_0 and stopping criterion is established.

Construct Ant Solutions. A set of m ants constructs candidate solutions to the problem. Each candidate solution is composed of solution components $c_i^j, (i = 1, \dots, L - 1)$, that are instantiations of binary random variables ($j \in \{0, 1\}$). A solution component c_i^1 indicates that threshold $t_k, (k = 1, \dots, K)$ is selected at i -th level of histogram, while c_i^0 indicates that it is not. Each solution component c_i^1 is associated with pheromone trail τ_i . A solution component c_i^1 is chosen with probability

$$p_i(c_i^1) = \frac{\tau_i}{\sum_j \tau_j}. \quad (1)$$

To meet the cardinality constraint of the number of thresholds (i.e. $\#(c_i^1) = K$), candidate solutions are constructed as follows. First, the probabilities $p_i(c_i), (i = 1, \dots, L - 1)$ are calculated for all components c_i^1 . Then, pseudo-random numbers $Y_i \in (0, 1), (i = 1, \dots, L - 1)$ corresponding to solution components are drawn from the standard uniform distribution. Finally, the random variable $Z_i = p_i - Y_i$ is calculated and K components corresponding to maximum realizations of Z_i are chosen.

Once all candidate solutions are constructed, they are evaluated according to the objective function. The value of best solution T_{bs} is stored in memory as best-so-far solution. T_{bs} can be updated in the subsequent iterations with the value of the solution obtained as current iteration best solution, if it has a better objective function value.

Apply Local Search. Once the best-so-far solution T_{bs} is established, it is further improved by applying local search based on the single-flip neighborhood. In the single-flip neighborhood a solution $T = (t_1, t_2, \dots, t_K)$ is a neighbor of T' if they are different in exactly one variable t . In addition, we assumed that the variable realizations in neighboring solutions must differ by a value of 1. This means that in case of M -level thresholding, $2M$ neighboring solution to T_{bs} is evaluated.

Update Pheromone Trails. The pheromone update rule is based on elitist strategy, which is popular ACO-variant and consists of two parts. First, a pheromone evaporation, which uniformly decreases all the pheromone values is performed. Second, the best-so-far solution from the current or earlier iteration is used to increase

the values of pheromone deposit on solution components that are parts of best solution. The update rule is given by:

$$\tau_{ij} \leftarrow (1 - rho) \tau_{ij} + wF(T_{bs}) \quad (2)$$

where $\rho \in (0, 1]$ is a parameter called evaporation rate, F stands for quality function and w is weighting parameter.

2.2 Honey Bee Mating Optimization

HBMO algorithm belongs to the general class of SI methods that models the behaviors of social insects. It is inspired by the marriage behavior of honey-bees [1, 5, 9]. A mating process starts with a dance performed by the queen who then flight far from the nest. Drones follow the queen and seven to twenty drones can mate with her. When the queen lays eggs, she fertilize some eggs by injecting at random a mixture of the sperms stored in the spermatheca.

In order to develop the algorithm, some simplifications was introduced to the model of mating procedure. Population of drones $D = \{D_1, D_2, \dots, D_{n_d}\}$ is generated once at the beginning of the algorithm and does not change during the algorithm. In contrast to the original HBMO algorithm it was decided to ignore the brood care role of workers during reproduction procedure. The mating flight starts with the queen Q initialized with some speed $S(0)$. The flight is continued until the spermatheca of capacity n_s is full or the speed drops below given threshold S_{min} and then the queen returns to the nest. After each mating iteration, the queen's speed is reduced according to the following formulae:

$$S(k+1) = \alpha S(k), \quad (3)$$

where $\alpha \in [0, 1]$ is the decreasing factor. Subsequent drones involved in the mating process are chosen randomly with uniform distribution from the set D . The probability of adding the sperm of selected drone D_{rand} to the spermatheca of queen Q (successful mating) is given by annealing function:

$$P(Q, D_{rand}) = e^{\frac{-|f(Q) - f(D_{rand})|}{S(k)}}, \quad (4)$$

where $f(Q)$ and $f(D_{rand})$ are the objective function values of the queen and the randomly chosen drone respectively. When the mating flight is over the queen starts laying n_c eggs and fertilize them by randomly selecting with uniform distribution a sperm from the queen's spermatheca. The brood B_j is generated by modifying the queen's genome using the following equation:

$$B_j = Q + \beta(P_{rand} - Q), \quad (5)$$

where $B_j = [B_j^1, B_j^2, \dots, B_j^c]$ is the j -th brood, P_{rand} is the randomly selected drone's genome from the spermatheca, $\beta \in [0, 1]$ is the randomly generated value with

uniform distribution during each breeding process and c is the size of the genome. In the next step, n_m randomly selected broods with uniform distribution are affected by the mutation defined by the following formulae:

$$B_{rand}^k = B_{rand}^k \pm (\delta - \varepsilon) B_{rand}^k, \quad (6)$$

where k and $\delta \in [0, 1]$ are randomly chosen with uniform distribution, $\varepsilon \in [0, 1]$ is predefined and the $+$ or $-$ sign occurs with equal probability. Finally the brood B_{best} with the highest value of objective function is selected. It replaces current queen if the objective function of B_{best} is higher than the objective function of the queen Q .

2.3 Firefly Algorithm

FA is an optimization technique developed in 2007 by Yang [22]. The algorithm is inspired by the flashing behavior of fireflies. Each firefly interacts with each other with certain strength. The attractiveness depends on the light intensity, which is associated with the objective function, and the distance between the fireflies. It is assumed that there is no sexual distinction between fireflies. Furthermore, less bright firefly will always be attracted by brighter one and the brightest firefly will be moved randomly.

The light intensity $I(r)$ varies according to the inverse square law

$$I(r) = \frac{I_s}{r^2}, \quad (7)$$

where I_s is the intensity at the source and r is the distance between two fireflies. Light is absorbed in the media with an absorption coefficient γ . The combination of the inverse square law and the absorption can be approximated to avoid singularity at $r = 0$ in I_s/r^2 by

$$I(r) = I_0 e^{-\gamma r^2}, \quad (8)$$

where I_0 is the original light intensity. As the firefly attractiveness is proportional to the brightness seen by its neighbor, the attractiveness function is determined by

$$\beta(r) = \beta_0 e^{-\gamma r^2}, \quad (9)$$

where β_0 is the attractiveness at $r = 0$. To determine the distance affecting the attractiveness between any two fireflies i and j at positions x_i and x_j Euclidean measure is used

$$r_{ij} = \|x_i - x_j\|. \quad (10)$$

In each iteration less attractive fireflies move to the brighter ones. The movement of the firefly i at location x_i to firefly j at location x_j is expressed by

$$x_i(t+1) = x_i(t) + \beta_0 e^{-\gamma r} (x_j - x_i(t)) + \alpha \varepsilon_i \quad (11)$$

where α is the randomization parameter, and ε_i is a vector of random numbers. In this work the vector ε_i is determined using a random number generator uniformly distributed in $[-0.5, 0.5]$.

3 Experimental Results

T All methods presented in this work were tested on real medical data. For this purpose, 450 images were collected from 50 patients (25 benign and 25 malignant). Each patient is represented by 9 images. The number of images was recommended by the specialists from the hospital and allows for correct diagnosis by a pathologist.

The cytological material was obtained by fine-needle biopsy from patients of the Regional Hospital in Zielona Góra, Poland. Biopsies without aspiration were performed under the control of ultrasonograph with a 0.5 mm diameter needle. Smears from the material were fixed in spray fixative (Cellfix by Shandon) and dyed with hematoxylin and eosin (H&E). The time between preparation of smears and their preservation in fixative never exceeded three seconds. All cancers were histologically confirmed and all patients with benign disease were either biopsied or followed for a year.

The images were recorded by a Sony CCD Iris color video camera mounted atop Axiophot microscope. The slides were projected into the camera with $160\times$ objective and $2.5\times$ ocular giving together an enlargement of $400\times$. Images are uncompressed TIFF files, 704×578 pixels, 8 bit/channel RGB. Sample images and thresholding results are presented in Fig. 1. To measure the *quality* of thresholding we applied the classification procedure described in Section 2. The three global optimization algorithms presented in Sections 2.1, 2.2, and 2.3 were involved to find optimal threshold values and compared for their performance (classification rate). To determine the optimal number of thresholds N^* the testing procedure was performed for $N = 2, \dots, 10$. The relationship between the number of thresholds and discriminative power of extracted features is shown in Fig. 2.

To compare the performance of the algorithms we use the number of objective function evaluations. Bearing in mind that all investigated methods are stochastic global optimization algorithms, we performed 10 replications for each algorithm and each number of thresholds. To keep the computational effort constant the length of each repetition is set to 500 objective function evaluations. The parameters of the algorithms have been adjusted in the process of preliminary testing and their values are as follows:

ACO population of ants = 10, $w = 2$, $\rho = 0.01$, $\tau_0 = 0.5$,
 HBMO population of drones = 100, $n_s = 50$, $n_c = 25$, $n_m = 3$, $\alpha = 0.98$, $\varepsilon = 0.5$,
 $S(0) = 1$, $S_{min} = 0.05$,
 FA population of fireflies = 20, $\gamma = 0.0001$, $\alpha = 10$.

Objective function values for the best found solutions and their median over 10 runs for all compared algorithms are presented in Table 1.

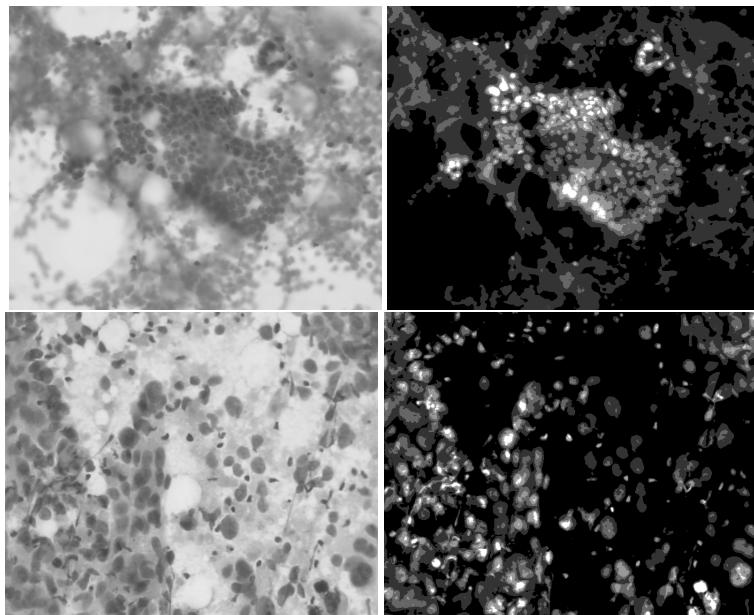


Fig. 1 Original images (left), and sample thresholding results for $N = 5$ (right)

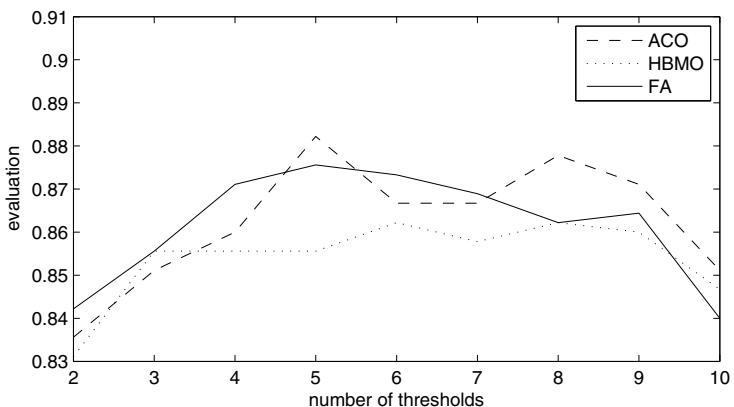


Fig. 2 The best achieved classification results depending on number of thresholds

Table 1 The best and median (in brackets) objective function values obtained from 10 runs for number of thresholds N from 2 to 10 using ACO, HBMO and FA algorithms

N	ACO	HBMO	FA 11T
2	0.8356 (0.8222)	0.8311 (0.8200)	0.8422 (0.8222) 11T
3	0.8511 (0.8433)	0.8556 (0.8422)	0.8556 (0.8311) 11T
4	0.8600 (0.8500)	0.8556 (0.8378)	0.8711 (0.8533) 11T
5	0.8822 (0.8589)	0.8556 (0.8411)	0.8756 (0.8667) 11T
6	0.8667 (0.8522)	0.8622 (0.8444)	0.8733 (0.8611) 11T
7	0.8667 (0.8456)	0.8578 (0.8356)	0.8689 (0.8378) 11T
8	0.8778 (0.8556)	0.8622 (0.8378)	0.8622 (0.8389) 11T
9	0.8711 (0.8511)	0.8600 (0.8411)	0.8644 (0.8356) 11T
10	0.8511 (0.8444)	0.8467 (0.8367)	0.8400 (0.8133) 11T

4 Conclusions

In the performed real-data experiment, the adjustment of segmentation parameters was driven by the feedback from classifier. The preliminary results indicated by high prediction accuracy seems promising. Application of kNN classifier allowed to find optimal number of thresholds, above which the results are worsening. The satisfactory explanation of this fact can be *the curse of dimensionality*. Increasing amount of irrelevant information has a serious influence on convergence of distances for any given pair of instances in feature space, what is reflected in k-NN classifier results. Further research should be performed in order to find better quality feature space. For example, the current feature set can be extended (or replaced) with another morphometric, topological or texture features.

Acknowledgements. The authors wish to thank Dr. Roman Monczak from the Regional Hospital in Zielona Góra, Poland for his great help and interesting discussions.

This research was partially supported by National Science Centre in Poland.

Paweł Filipczuk is a scholar within Sub-measure 8.2.2 Regional Innovation Strategies, Measure 8.2 Transfer of knowledge, Priority VIII Regional human resources for the economy Human Capital Operational Programme co-financed by European Social Fund and state budget.



HUMAN CAPITAL
HUMAN – BEST INVESTMENT!



Lubuskie
Worth your while

EUROPEAN UNION
EUROPEAN
SOCIAL FUND



References

1. Abbasss, H.A.: Marriage in honey bee optimization (hbo): a haplometrosis polygynous swarming approach. In: Proc. Cong. on Evolutionary Computation, pp. 207–214 (2001)
2. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Trans. on Information Theory 13(1), 21–27 (1967)
3. Devijver, P.A., Kittler, J.: Pattern Recognition: A Statistical Approach. Prentice-Hall, London (1982)
4. Dorigo, M., Stuetzle, T.: Ant colony optimization: Overview and recent advances. In: Gendreau, M., Potvin, J.-Y. (eds.) Handbook of Metaheuristics. International Series in Operations Research & Management Science, vol. 146, pp. 227–263. Springer (2010)
5. Fathian, M., Amiri, B., Maroosi, A.: Application of honey bee mating optimization algorithm on clustering. Applied Mathematics and Computation 190, 1502–1513 (2007)
6. Filipczuk, P., Kowal, M., Obuchowicz, A.: Automatic breast cancer diagnosis based on K-means clustering and adaptive thresholding hybrid segmentation. In: Choraś, R.S. (ed.) Image Processing and Communications Challenges 3. AISC, vol. 102, pp. 295–302. Springer, Heidelberg (2011)
7. Filipczuk, P., Kowal, M., Obuchowicz, A.: Breast fibroadenoma automatic detection using k-means based hybrid segmentation method. In: Proc. IEEE International Symposium on Biomedical Imaging: from nano to macro, ISBI 2012, Barcelona, Spain, pp. 1623–1626 (2012)
8. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall, New Jersey (2001)
9. Horng, M.-H.: A multilevel image thresholding using the honey bee mating optimization. Applied Mathematics and Computation 215, 3302–3310 (2010)
10. Jeleń, L., Fevens, T., Krzyżak, A.: Classification of breast cancer malignancy using cytological images of fine needle aspiration biopsies. Int. J. Appl. Math and Comp. Sci. 18(1), 75–83 (2008)
11. Kapur, J.N., Sahoo, P.K., Wong, A.K.C.: A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision Graphics Image Processing 29, 273–285 (1985)
12. Kohavi, R., John, G.: Wrappers for feature subset selection. Artificial Intelligence Journal 97, 273–324 (1997)
13. Kowal, M., Filipczuk, P., Obuchowicz, A., Korbicz, J.: Computer-aided diagnosis of breast cancer using gaussian mixture cytological image segmentation. J. of Medical Informatics & Technologies 17, 257–262 (2011)
14. Malek, J., Sebri, A., Mabrouk, S., Torki, K., Tourki, R.: Automated breast cancer diagnosis based on GVF-Snake segmentation, wavelet features extraction and fuzzy classification. J. of Signal Processing Systems 55, 49–66 (2009)
15. Marciniak, A., Obuchowicz, A., Monczak, R., Kolodziński, M.: Cytomorphometry of Fine Needle Biopsy Material from the Breast Cancer. In: Proc. 4th Int. Conf. on Computer Recognition Systems CORES 2005, pp. 603–609. Springer (2005)
16. Niwas, I.S., Palanisamy, P., Sujathan, K.: Wavelet based feature extraction method for breast cancer cytology images. In: Proc. IEEE Symp. on Industrial Electronics & Applications, ISIEA, pp. 686–690 (2010)
17. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Sys. Man. and Cyber. 9, 62–66 (1979)
18. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. J. Electronic Imaging 13(1), 146–165 (2003)

19. Shapiro, L.G., Stockman, G.C.: Computer Vision. Prentice Hall, New Jersey (2002)
20. Soltani, M., Chaari, A., Hmida, F.B.: A novel fuzzy c-regression model algorithm using a new error measure and particle swarm optimization. *Int. J. Appl. Math and Comp. Sci.* 22(3), 617–628 (2012)
21. Würflinger, T., Stockhausen, J., Meyer-Ebrecht, D., Böcking, A.: Robust automatic coregistration, segmentation, and classification of cell nuclei in multimodal cytopathological microscopic images. *Computerized Medical Imaging and Graphics* 28, 87–98 (2004)
22. Yang, X.S.: Nature-Inspired Metaheuristic Algorithms. Luniver Press, Frome (2010)

Detecting Overlapped Nuclei Regions in the Feulgen-Stained Cytological Smears

Bogusław D. Piętka and Annamonika Dulewicz

Abstract. Projects related to clinical implementation of computerized image processing in cytology face a common problem of distinguishing between artifacts and the objects of interest which should be measured and analyzed. Secondary Screening Instruments are computerized devices used to detect possible presence of cancerous or precancerous cells and, independently or in collaboration with pathologist, select slides for additional manual review. In order to do the work efficiently they have to detect actual abnormalities distinguishing them from artifacts in the form of overlapping cells or nuclei. The paper reports a trial approach to perform this important discrimination.

1 Introduction

The potential benefit of Secondary Screening Instruments (SSI) is that they increase the overall sensitivity of a cytological screening by detecting the presence of possible abnormalities missed during a primary screening. Because primary screening (i.e. first time examination of a slide) is not 100

Most projects related to clinical implementation of computerized image processing in cytology face a common problem of distinguishing between artifacts and the objects of interest which should be measured and analyzed. Since most of algorithms for cancer cell identification rely on some kind of abnormality detection, artifacts left in a sample would generate too many undesired, false-positive alarms, making such a system impractical.

Bogusław D. Piętka · Annamonika Dulewicz
Laboratory of Fundamentals of Computer-Aided Image Diagnostics,
Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences,
ul. Ks. Trojdena 4, 02-109 Warsaw, Poland
e-mail: bpietka@ibib.waw.pl

Generally, artifacts are undesired objects or phenomena influencing the appearance of a smear and obstructing, or even preventing, proper analysis of important factors of cytological sample. Several examples of the artifacts are presented in Fig. 1.1.

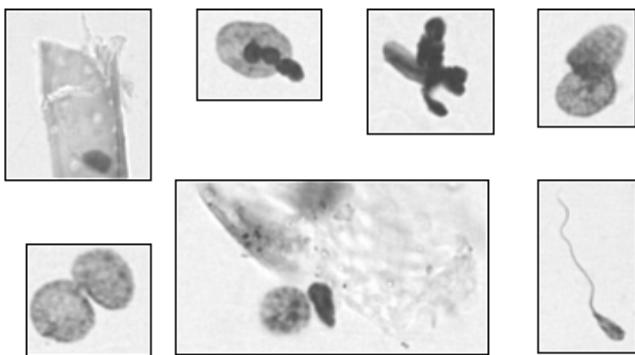


Fig. 1.1 Examples of the artifacts found in a cytological smear (Feulgen-stained urine sample)

Although there are different types of artifacts, in this paper we are focusing on just one, particular sort of them; two overlapping nuclei (Fig. 1.2).



Fig. 1.2 Examples of particular type of artifacts in the form of two normal (non-cancerous) overlapping nuclei

2 Searching for Textural Markers of the Overlap

In the case of evident nuclei overlap it is relatively easy for a human observer to distinguish two regions of different textural nature (Fig. 2.1). Let us try to find some objective, quantitative indicators of the occlusion. There are three basic methods used to describe the texture of a region: statistical, structural and spectral. Since the structural and spectral techniques are primarily used to find a kind of periodicity, directionality or repetitive pixel arrangement, they are not well suited for the problem under consideration. Therefore, the only promising approach seems to be statistical one. Starting from the simplest, first order measures of a region texture,

the his-tograms $h(z_i)$ ($i = 1, 2, \dots, N$, $z_i = 0, 1, \dots, N-1$, $N = 256$) of the two distinct areas ($A \cup B$ and C) from figure 2.1 are computed (Fig.2.2). Actually, prior to histogram calculation, all images are normalized by means of the contrast stretching to saturate the bright background (255).

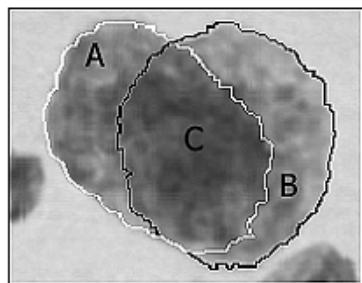


Fig. 2.1 An example of manually selected regions used for studying the overlap phenomenon

In general, intensity distributions of regions A and B are computed sepa-rately. Since our experiments proved they are highly similar, it was decided to combine them into one region. Thus, we get two histograms describing one, overlapping object:

$$H_{A \cup B} = H_A + H_B \text{ and } H_C$$

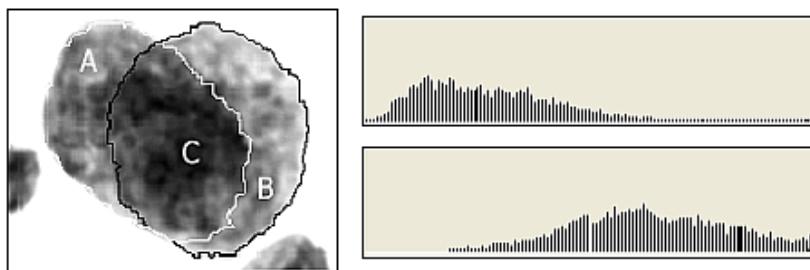


Fig. 2.2 The image from Fig.2.1 after normalization and grey level distributions of two distinct areas: overlapped (C-top) and non-overlapped ($A \cup B$ -bottom)

Although differences between locations of the two histograms on the grey level axes are clearly visible, they do not constitute sufficient overlap indica-tion. They are not even the most essential for the overlap identification. For example, a single malignant nucleus also tends to have its histogram shifted to the darker grey levels as a result of increased presence of heterochromatin (Fig. 2.3). However, we have noticed statistical significance of the higher order central moments of the histograms in distinguishing such cases from the overlaps.

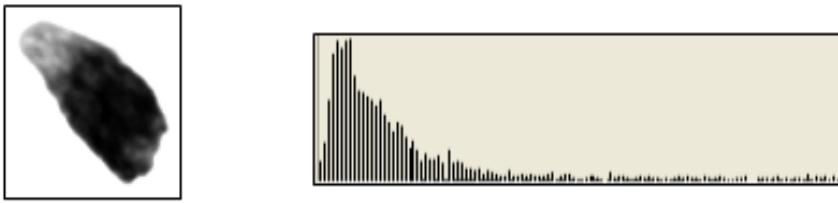


Fig. 2.3 Suspicious nucleus and the grey-level distribution of its dark region

To construct objective statistical descriptors of a texture, four parameters of the grey-level distribution were computed. In addition to the mean value μ they were: variance, skewness and kurtosis according to the following formulas:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N [h(z_i) - \mu]^2$$

$$\gamma = \frac{1}{(N-1) \cdot \sigma^3} \sum_{i=1}^N [h(z_i) - \mu]^3$$

$$\delta = \frac{1}{(N-1) \cdot \sigma^4} \sum_{i=1}^N [h(z_i) - \mu]^4$$

where the mean value μ is defined as

$$\mu = \frac{1}{N} \sum_{i=1}^N [h(z_i) \cdot z_i]$$

and $N = 256$ is the total number of grey levels in the histogram.

Fifty image samples for each of the three area types (normal, dark-malignant and overlapped) were collected, processed and the resulting data supplied to unsupervised clustering algorithm. Our goal was not to perform precise discriminant analysis but to discover any general patterns that may be present in the collected data set.

3 Experimental Results

In the very beginning of the data analysis process it was found that the mean value of the histograms has the lowest discrimination power in comparison with others. Therefore, it was decided to perform unsupervised clustering of the data basing only on the three descriptors: variance, skewness and kurto-sis. Moreover, since the goal of the investigations was to get some qualitative knowledge about the phenomenon

we are going to show the results graphically. For this purpose the KlustaWin software package was used that is able to perform different kinds of data clustering and present the results as a two or three-dimensional scatter-plots [5]. The software is an open-source project for scientific purposes.

It may be informative to see the scatter-plots from three different perspectives (Fig. 3.1-3.3) where selected histogram parameters were assigned to the X, Y, Z axes as follows

$$\sigma^2 \implies X \text{ (variance)}$$

$$\gamma \implies Y \text{ (skewness)}$$

$$\delta \implies Z \text{ (kurtosis)}$$

Numerical values of the data on every axis were normalized to the same range 0-1.

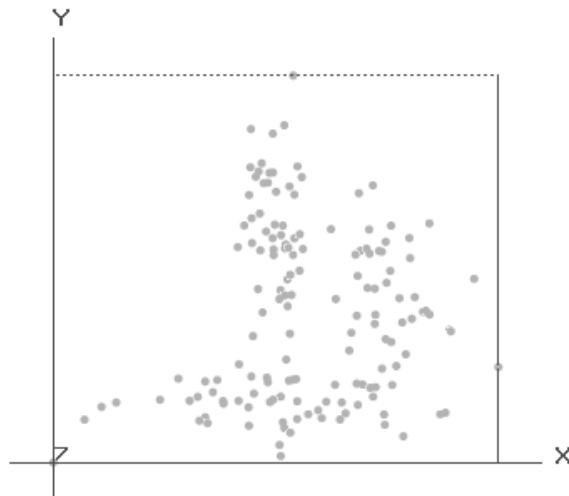


Fig. 3.1 2-D scatter-plot as can be seen from variance-skewness surface perspective

From the perspective of variance-skewness surface it is impossible to clearly distinguish any clusters. However, perceptive observer would draw a division line orthogonal to X axis at approximately $x=0.55$. The zones created in this way represent overlapping textures (left) and non-overlapped regions (right).

It may be easily noted that the scatter-plot from Fig. 3.2 (variance-kurtosis) demonstrates much better separation than that from Fig. 3.1 (variance-skewness), and mainly thanks to the kurtosis (Z axis). This surface is best suited for distinguishing between dark-malignant regions and overlapped regions which usually have high density as well. Finally, the last scatter-plot in Fig. 3.3 is definitely the best one. It seems that the pair of descriptors skewness-kurtosis would be enough to

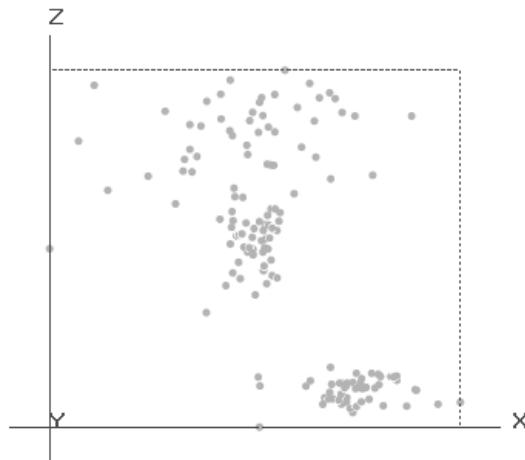


Fig. 3.2 2-D scatter-plot as can be seen from variance-kurtosis surface perspective

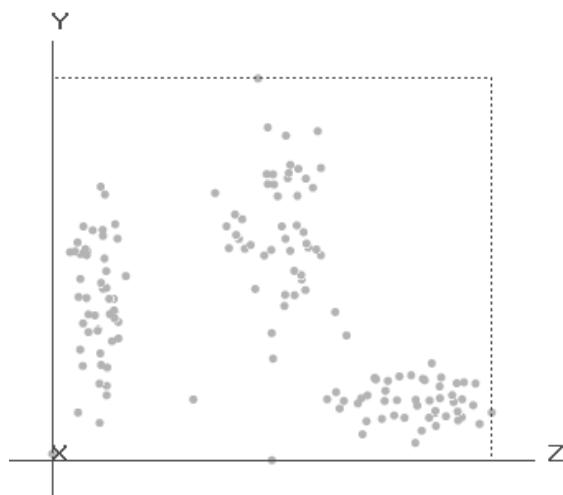


Fig. 3.3 2-D scatter-plot as can be seen from kurtosis-skewness surface perspective

obtain very good separation between the three kinds of textures under study. Suspicious textures are characterized by the high kurtosis (cluster on the right). On the other hand, non-overlapped textures histograms have wide tails implicating low values of the kurtosis. To get an impression of the data distribution in full three-dimensional space there is a 3-D scatter-plot shown in Fig. 3.4 below.

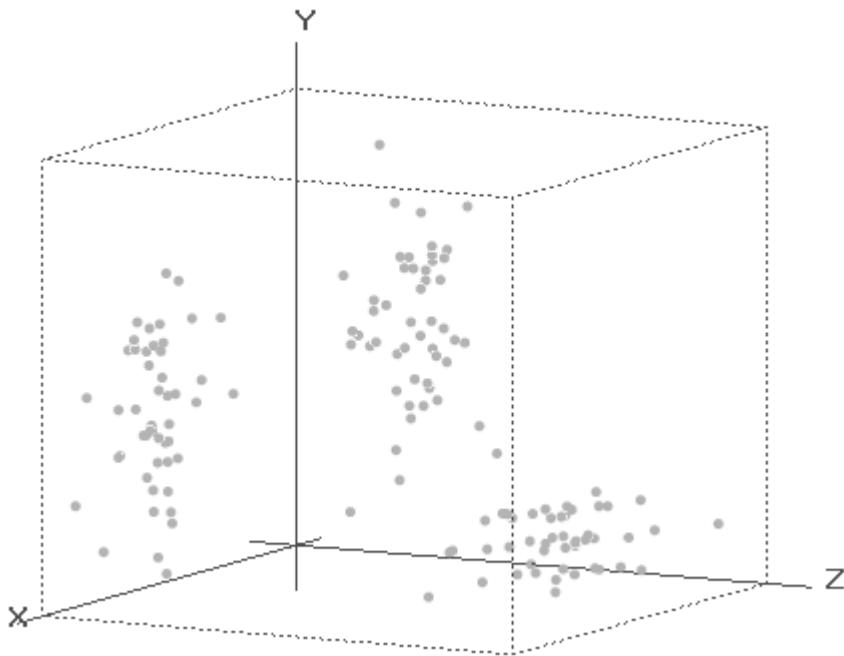


Fig. 3.4 3-D scatter-plot combining all three parameters; variance, skewness and kurtosis

4 Summary

It was found useful when describing textural features of nuclear overlaps to divide the problem into two distinct levels of complexity. First, in every initially extracted object, a trial is performed to divide the total area into two regions with different textural properties. If the division may be performed successfully, a detailed, histogram-based analysis of the textures is conducted to identify overlapping regions.

Since it was just an initial approach to solve the problem, the regions were extracted manually and then subjected to computer analysis by means of image processing tools. It was shown that combination of properly selected descriptors of the region texture (skewness and kurtosis) can perform quite good separation in the data space. Simple reclassification based on the centroids of the clusters (Fig. 3.4) is presented in the table below (Fig. 4.1).

Unfortunately, at this particular moment we have not finalized the task of designing a fast and reliable procedure for automatic segmentation of the whole object which is necessary to complete the task. The work is under progress and will be the subject of a paper in near future.

Expert →	Non-Overlapped	Overlapped	Cancerous
Non-overlapped	50	3	0
Overlapped	0	45	0
Cancerous	0	2	50

Fig. 4.1 Results of the sample textures reclassification (150 items)

References

1. Ravishankar, A., Lohse, G.: Identifying high level features of texture perception. CVGIP: Graph. Models Image Process. 55(3), 218–233 (1993)
2. Haralick, R.M.: Statistical and Structural Approaches to Texture. In: van Leeuwen, J. (ed.) Computer Science Today. LNCS, vol. 1000, pp. 786–804. Springer, Heidelberg (1995)
3. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics SMC-3(6), 610–620 (1973)
4. Harris, K.: (Rutgers University, USA), Heitler B. (University of St Andrews, UK), KlustaWin - A program for unsupervised classification of multidimensional continuous data, <http://www.st-andrews.ac.uk/~wjh/klustawin/>

Density Invariant Detection of Osteoporosis Using Growing Neural Gas*

Igor T. Podolak and Stanisław K. Jastrzębski

Abstract. We present a method for osteoporosis detection using graph representations obtained running a Growing Neural Gas machine learning algorithm on X-ray bone images. The GNG induced graph, being dependent on density, represents well the features which may be in part responsible for the illness. The graph connects well dense bone regions, making it possible to subdivide the whole image into regions. It is interesting to note, that these regions in bones, whose extraction might make it easier to detect the illness, correspond to some graph theoretic notions. In the paper, some invariants based on these graph theoretic notions, are proposed and if used with a machine classification method, e.g. a neural network, will make it possible to help recognize images of bones of ill persons. This graph theoretic approach is novel in this area. It helps to separate solution from the actual physical properties. The paper gives the proposed indices definitions and shows a classification based on them as input attributes.

1 Introduction

Osteoporosis is now a frequent illness. It strikes mainly female population. Treatment is long, usually life-long, with abnormally high costs. An early diagnosis is of the utmost value. Due to aging society, osteoporosis is becoming more and more prevalent illness.

Several methods, which are based on X-ray images, micro-tomography, and some which are invasive. The main objective of this paper is to propose a novel approach to the analysis of tomography images which computes some invariants based

Igor T. Podolak · Stanisław K. Jastrzębski
Institute of Computer Science, Jagiellonian University
e-mail: {igor.podolak, stanislaw.jastrzebski}@uj.edu.pl

* This work was in part supported by the NCN grant N N519 654 480.

on a graph representations of 3-D images. This approach uses first a Growing Neural Gas (GNG) machine learning algorithm to obtain the graph, then graph-theoretic methods to compute the invariants. In the end, other machine learning methods are used to classify images.

1.1 *Background and Motivation*

Osteoporosis is defined as “a systematic skeletal disease characterized by low bone mass, deterioration of trabecular architecture and increased fragility of bone” [13]. Main methods to assess the fracture risk are based on densitometric measurements: Bone Mineral Density and Content (BMD, BMC) since the decrease of bone mass is the most frequent observation.

In a density measurement we find a deviation from population mean of the bone density. However, this method it is feasible to detect advanced osteoporosis, because in early stages of disease density does not decrease noticeably. Even though decreased BMD is said to be the primary symptom, very frequently patients with normal BMD have their bones fractured. Early diagnosis of is crucial, since changes in bone structure are often irreversible.

The current, image recognition based, methods compute several factors which may help to quantify the bone structure. Among them node-number, node-to-node strut count, terminus-to-terminus strut count, node to terminus ratio, trabecular bone pattern factor, structure model index, structural anisotropy measures, and several others. Most are based on the image recognition using segmentation, dilation, and other similar methods (see, e.g. [7, 8, 11]). Recently, topology based methods are also employed [1].

1.2 *Contributions and Organization of This Paper*

The main objective of this paper is to suggest an alternative approach to the problem, which may yield new method for detection early bone changes. It is in part a graph theoretic approach (as, e.g., suggested in [12]), but we propose here a novel method to obtain such a graph using a machine learning methodology. We believe that the proposed approach will provide methods to assess risk of osteoporosis earlier than it is possible now.

Most of the methods used now try to extract, from images, some physical values, which are then used for a classification process. We want to find a solution which is based on more *abstract* and holistic measures of the bone structure. Using GNG, a graph representing bone is obtained, and then some theoretical measures of it are found. Since this is a different approach, we believe that if it shall be added to the current systems, a substantial growth of classification accuracy will be obtained. It is even more significant, because it is not clear what are the main factors indicating early osteoporosis. In the long run our goal is to build detector of early changes in

bone, and the first step is building density invariant classifier of bones afflicted with osteoporosis.

In the work below first present how the extraction of graph is performed, then describe all the proposed invariants, then present our experiments and the future work.

2 Preliminaries

It is important to make the basic distinction in the structure of bone, which consists of shell and inner part, both responsible for its strength and endurance. We have taken bone density measurements using X-Ray scan, as in Fig. 1, where brightness of the pixel reflects its density measured in g/dm^3 . We shall investigate the topological structure of the bone, which agree with its physical properties. X-ray scans are noisy, and it may not be easy to extract 3D information from them.

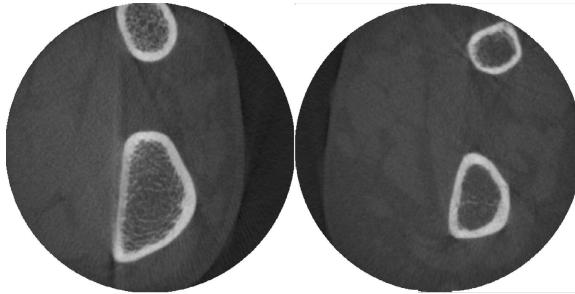


Fig. 1 An X-Ray of a healthy and a bone with osteoporotic changes

2.1 Growing Neural Gas Algorithm

For this purpose we have used Growing Neural Gas algorithm, an unsupervised clustering algorithm proposed by Fritzke [5]. Let some process \mathcal{P} generate points in K -dimensional space. A GNG model is a graph composed of nodes, each with a reference vector w in the K -dimensional space and an *error* attribute, and edges which connect some nodes. Each edge has an *age* attribute. For each data x , the algorithm finds the first n_f and second n_s nearest nodes. All nodes emanating from n_f have their *age* increased. If the nodes are not connected, an edge is added, otherwise it's *age* is set to 0. Edges with *age* higher than some maximum set, are removed. Nodes n_f and n_s are moved towards x by ε_f and ε_s fractions, respectively. Error of n_f is increased by factor $\|x - w_f\|^2$.

Every given number of epochs a node n_q with highest error is found, and a new node is added between n_q and one of its connected neighbors with highest error. This corresponds to adding nodes in areas of high density. Using such procedure, a GNG clusters the data space into separate clusters described with connected components.

Using GNG is justified by its theoretical properties which were crucial for our research. To name the most important, the graph resulting from GNG is in strong correspondence with original object's physical strength [6]. Therefore, the resulting graph reflects density distribution and physical links between bone elements. The resulting graph is spatial, because it also carries position information. Even more important, every point in our 3D space also possesses density attribute, therefore density of a vertex in GNG graph will be equal to weighted average density of points in its neighborhood. Let $G = (V, E)$ denote graph output from GNG. Now ϕ_v is a vertex density and $v_{x/y/z}$ is its position in X-ray space.

It was crucial to produce high-resolution graphs in order to capture subtle holes and irregularities in bone structure. Unfortunately, original GNG implementation was much too slow to meet our aims. We have implemented GNG in C++ with Rcpp port to R, using very fast implementation suggested by Fiser [3]. Procedures responsible for searching for the closest nodes, and biggest error were the bottlenecks of the original implementation. Thanks to Uniform Grid and Lazy Heap structures, we can bring down running time of one query to $O(1)$. This implementation scales very good with the number of nodes, therefore we managed to produce high-resolution graphs.

3 Methodology

Before diving into the details, we shall give some intuition behind the research. Everything in nature that is in general sense symmetrical and regular, is good. We have assumed that throughout all healthy bones, there are maintained invariant of density and shape properties, like regularity, number of components, etc. Our experiments have proven this assumption, see Fig. 2. All bones, that we have examined, share some general way of organization. E.g., it is well known that healthy bone structure is anisotropic (see e.g. [10]).

First the exterior and interior parts are extracted, then processed separately, since they have significantly different biological structure. Afterwards 3D points were fed into GNG algorithm sampled accordingly to density measured by X-Ray. Then a set of graph measures selected for this problem are computed. These are the indices which we claim to be useful for later ill/healthy classification. Using these indices, a training set is built, which is later used with some machine learning approach. To prove that these indices are invariant of densities, we have standardized bone densities of healthy and ill patients. Different different classifiers were used at the end.



Fig. 2 GNG graphs of an ill and a healthy bone — notice changes in structure

3.1 Extraction

Since denser points will be more significant for GNG, we need to proceed very carefully with extraction not to include very dense shell areas in the inner part of the bone. The objective is to get rid of X-ray noise and separate, with strong confidence, bone into data sets (see Fig. 3). To achieve the separation, we fit a two Gaussian distributions sum model using Expectation–Maximization EM method to obtain a rough separation.

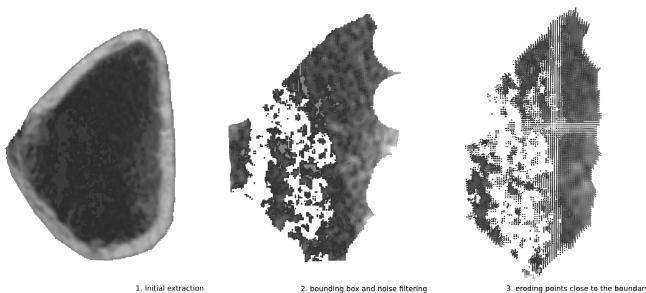


Fig. 3 Extraction steps described in the text

In the second step generalized convex hull is fit to the inner part, with points close to the edge removed. They can be dense hindering the exposure of more interesting inner substructure and holes by GNG. The last step consists of fitting Gaussian distribution to each coordinate and density and removing unlikely points — in fact we compute a bounding box for the bone in four dimensions. Processing of the shell part follows roughly the same schema.

GNG parameters were fitted experimentally to obtain fast convergence. Actually, the graphs obtained proved to be highly independent of parameter values. Along with the spatial information about a point, we attach probability of its occurrence. One could use linear scaling to range [0,1], however it's use results in a healthy

person graph being poorly connected, because of very dense parts that would have very high probability of occurrence. Having this in mind, we used scaled sigmoidal function

$$p(x) = \frac{1}{1 + e^{-\beta x}}, \quad \beta = \frac{1}{threshold} \ln \left(\frac{\alpha}{1 - \alpha} \right), \quad (1)$$

where β is fitted to maintain invariant that 70% most dense points have got α or bigger probability function. Threshold is found by fitting density function to points' density histogram, while α is picked during the experiment. Points are sampled by GNG algorithm according to this probability.

In the next step, using the graph representation, we compute some quantities, that in our view describe well the topological structure of GNG graph.

3.1.1 Interior/Exterior Densities

Because we cannot make indices completely independent of density, we feed into the classifier densities of the bone. Hopefully, it will be able to learn to scale other indices accordingly.

3.1.2 Graph betweenness and Graph Communities

Healthy bone structures are much more regular. A graph's edge represents the likeliness that two vertices represent the same cluster in a bone. Therefore, densely connected sub-graphs represent separate biological substructures of a bone. Thus, we want to measure the *number of bridges* (i.e. edges whose removal increases the number of connected components) in the graph, which is correlated with the number of dense substructures in the bone.

Naturally, bridges are very unlikely to occur in a densely connected GNG, therefore we measure *generalized bridges*, that is edges with high number of shortest paths that cross them, i.e. the edge *betweenness*, see Fig. 4. Similarly we can find

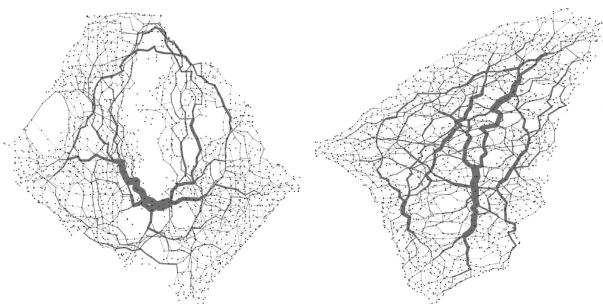


Fig. 4 Edge betweenness visualization of an osteoporotic bone (on the left) and a healthy one (on the right). These are the most important edges in the bone, also revealing the underlying structure of the bone. This structure differs noticeably in an ill bone

number of graph communities, i.e. densely connected sub-graphs. We have used `fastgreedy.community()` procedure from the `igraph` package in R [9]. *Community membership*, see Fig. 5, induces partition of the graph. Both measures are strongly correlated with the bone overall health, whereas are not affected by density, but rather by density distributions.

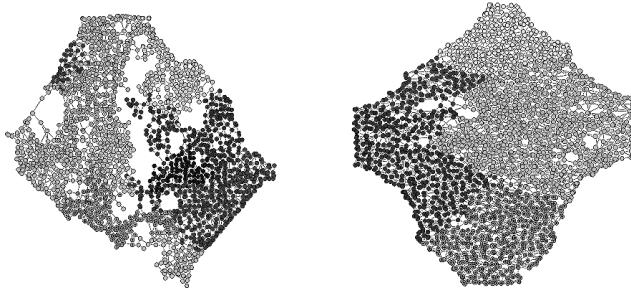


Fig. 5 Communities found for an osteoporotic and a healthy bone

3.1.3 Generalized Vertex Degree

The most basic graph measure is the degree sequence. GNG actually performs a bone triangulation, therefore we need a more generalized version of vertex degree to overcome this difficulty. Let $d_{u,v}$ denote Euclidean distance of two nodes. Let R be a constant radius of sampling sphere. We define

$$B_v = \{u : d_{u,v} \leq R\}, \quad (2)$$

and now we may define a *generalized vertex degree* as $\hat{d}_v = \#B_v$.

3.1.4 Ball Density

Similarly, we compute *distribution of ball density*. We take the defined above B_v and treat neurons found in this ball as "samples" of density of the ball of radius R . We fit a cube into our ball, and partition it into 27 identical cubes ($3 \times 3 \times 3$). Let c be the center point of a cube. For B_v cube density (see Fig. 6) is defined as

$$\phi_{subcube} = \frac{\sum_{u \in subcube} \phi_u * \exp \frac{-d_{u,c}}{R}}{\sum_{u \in subcube} \exp \frac{d_{u,c}}{R}}. \quad (3)$$

Now the *generalized density* is defined as

$$\hat{\phi}_u = \frac{1}{27} \sum_{subcube \in cube} \phi_{subcube}. \quad (4)$$

Let μ denote mean of $\phi_{subcube}$. We also define h_u – ball homogeneity as

$$h_u = \max ((\phi_{subcube} - \mu)^2). \quad (5)$$

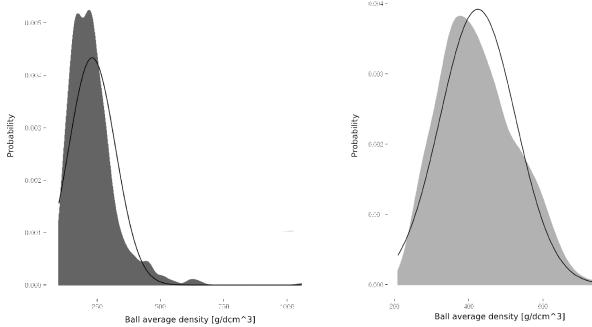


Fig. 6 Generalized ball density distribution in healthy and ill bones. Ill bones have skewed distributions due to higher number of very dense regions

3.1.5 Opposite Nodes Path Lengths

For the inner part slices we find nodes which lie on the border and then find the shortest paths between opposite nodes. This can be quickly computed using the `igraph` package functions. It appears that the mean shortest path for healthy bones is lower with lower variance too. This comes from the fact, that ill bones have lower homogeneity.

4 Experiments and Results

We have processed and extracted around 2000 X–Ray scans, approximately 200 per person. One measurement consisted of randomly picked 15 scans formed into a 3D point cloud. Using randomized α parameter and different density scaling we obtained almost 700 samples for machine learning algorithm.

Each training example consisted of a vector of the following values, all computed from graphs (all normalized): density means and variances (both inner and outer parts), bridge number means and variances (inner/outer), number of communities (inner/outer), ball density mean and variance (inner), ball homogeneity and skewness (inner), generalized vertex degree (inner) opposite nodes mean path lengths and variances (inner). Some of them mirror the bone’s density, some show the graphs connectivity, some are based on the paths found in the graph.

Edge weights are one of the cornerstones of the methodology described in this paper because, as it can be seen from the theoretical properties of GNG graphs,

short edges are likely to connect nodes belonging to some common dense regions. In indices like edge betweenness, communities, or ball density, edges play a major role.

We have used vectors of computed indices as an input for a learning algorithm. To compare generalization, we used different classification schemes: feed-forward neural networks, SVM's, decision trees and random forest, see Tab. 1. To assess accurately the generalization error, we used the Efron's “.632 estimator” [2], computing each experiment for 100 bootstrap experiments. This approach pulls down the optimism of a cross-validation.

Table 1 Experiment results

Training algorithm		$Err^{(.632)}$ generalization error
nnet	feed-forward neural network	0.096
RSNNS	feed-forward neural network	0.080
Support Vector Machine	radial kernel	0.140
	linear kernel	0.150
rpart	decision trees	0.160
random forest		0.150

It must be noted, that these are results which completely do not use any densitometry. Actually, the bone densities in the experiment were artificially standardized, so that the system would use rather the GNG structure description, not actual density values. The actual osteoporosis detection systems arrive at a 92–95% detection rate. Our system is not far behind.

5 Future Work

It is a report from an ongoing work. The main objective is to find measures which will make it possible to predict whether a person, while still healthy, may be susceptible to some form of osteoporosis. The now used methods are able to detect the illness only when it is already at a fully developed stage. The possible treatment is then only symptomatic.

We believe that the proposed approach, if used together with current methods, will pull up the correctness rate of osteoporosis detection. We want to extend this work in the direction of detecting people's liability to this illness, long before the actual symptoms become to be seen.

The authors wish to express their thanks to P. Dlotko and Z. Tabor for making their image databases available to us.

References

1. Drótko, P.: Private Communications (2012)
2. Efron, B.: Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7, 1–26 (1979)
3. Fiser, D., Feigl, J., Kulich, M.: Growing Neural Gas Efficiently. *Neurocomputing* (2012) (in press)
4. Fritzke, B.: Kohonen Feature Maps and Growing Cell Structures a Performance Comparison. *Advances in Neural Information Processing Systems* 5, 115–122 (1993)
5. Fritzke, B.: A Growing Neural Gas network learns topologies. *Advances in Neural Information Processing Systems* 7, 625–632 (1995)
6. Fritzke, B.: A self-organizing network that can follow non-stationary distributions. In: *International Conference on Artificial Neural Networks*, Lausanne, pp. 613–618 (1997)
7. Garrahan, N.J., Mellish, R.W.E., Compston, J.E.: A new method for the two-dimensional analysis of bone structure in human iliac crest biopsies. *J. Microsc.* 142, 341–349 (1986)
8. Hahn, M., Vogel, M., Pompeius-Kempa, M., Delling, G.: Trabecular bone pattern factor – a new parameter for simple quantification of bone micro architecture. *Bone* 13, 327–330 (1992)
9. R: A language and environment for statistical computing. R Foundation for Statistical Computing (2005)
10. Saha, P.K., Wehrli, F.W.: A robust method for measuring trabecular bone orientation anisotropy at in vivo resolution using tensor scale. *Pattern Recognition* 37, 1935–1944 (2004)
11. Tabor, Z., Rokita, E.: Quantifying deterioration of bone tissue from grey-level images. *Md. Eng. Phys.* 29, 497–504 (2007)
12. Tabor, Z.: Detecting surfaces of minimal cut – a graph theoretical approach. *Inżynieria Materiałowa* 4, 439–442 (2008)
13. World Health Organization, Assessment of fracture risk and its application to screening for post-menopausal osteoporosis, Technical Report Series 843 (1994)

Cost Sensitive Hierarchical Classifiers for Non-invasive Recognition of Liver Fibrosis Stage

Bartosz Krawczyk, Michał Woźniak, Tomasz Orczyk, and Piotr Porwik

Abstract. Liver Fibrosis caused by the Hepatitis Virus type C (HCV) may be a serious life-threatening condition if it is not diagnosed and treated on time. Our previous research proved that it is possible to estimate liver fibrosis stage in patients with diagnosed HCV only using blood tests. The aim of our research is to find a safe and non-invasive but also inexpensive diagnostic method. As not all blood tests are equally expensive (not only in meaning of money, but also time of analysis), this article introduces a Cost Factor to the hierarchical classifiers. Our classifier has been based on a C4.5 decision tree building algorithm enhanced with a modified EG2 algorithm for maintaining a cost limit.

Keywords: machine learning, cost sensitive classification, hierarchical classifier, decision trees, multi-stage pattern recognition, medical informatics, liver fibrosis.

1 Introduction

According to the Polish National Institute of Hygiene about 730 thousand people living in Poland are infected with the Hepatitis Virus type C. This liver disease often leads to a condition called liver fibrosis, which in its final stage may cause a liver cirrhosis - a serious life-threatening condition. Hepatitis in most cases has a chronic form, without an inflammation or any other visible symptoms.

For an adequate treatment it is important to estimate the degradation of the infected organ. The liver biopsy is often used as a diagnostic method of choice, but it is

Bartosz Krawczyk · Michał Woźniak
Department of Systems and Computer Networks, Wrocław University of Technology,
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: {bartosz.krawczyk, michał.wozniak}@pwr.wroc.pl

Tomasz Orczyk · Piotr Porwik
University of Silesia, Institute of Computer Science, Bedzinska 39,
41-200 Sosnowiec, Poland
e-mail: {tomasz.orczyk, piotr.porwik}@us.edu.pl

an invasive, potentially painful and even life-threatening procedure (and because of that it cannot be repeated in short periods of time). There are also some alternative, non invasive tests, which can be divided between biological and physical [5]. In the first group there are tests like ELF® [6, 7] or FibroTest® [3], but they are expensive commercial products and are not refunded by the Polish National Health Fund. The second group mainly consists of liver elastography. All the alternative liver fibrosis diagnostic methods use biopsy as the reference, but it is important to remember that also this method can give false results in up to 35% of cases [1, 15].

In our previous works we have proven that it is possible to estimate liver fibrosis stage, basing on a non specific blood tests [8, 10]. The main advantages over the biopsy of this diagnostic method is the fact it does not require hospitalisation and it can be repeated in regular periods of time without any damage to the patient. Due to limited funds for patients examination the cost of examination is also an important factor, so we have decided to introduce a cost factor into our classifiers. Our goal is to minimize the amount of expensive and time consuming blood tests without a noticeable raise of classification error rate. This would ease the doctors' decision to send a patient for such examination regularly.

From a technical point of view we have used a multi-stage hierarchical classifier in the form of decision trees, built using a C4.5 algorithm. For the purpose of including a complex cost limiting functionality we have used a modified EG2 formula.

2 Liver Fibrosis Recognition

Liver fibrosis is a condition where fibrous tissue accumulates in the liver. The stage of fibrosis can be determined by histological examination of liver tissue sample. The result may be classified according to different scales, but in this article the METAVIR [2] classification is used. The METAVIR has been specifically designed and validated for patients with hepatitis C. Accuracy of all classification systems relying on a histological image of the liver depend on a sample size and doctor's experience.

For the presented research we acquired a medical data records from 103 patients of the Gastroenterology and Hepatology Branch of the Independent Public Central Hospital of the Silesian Medical University.

We have also composed a cost factor basing on an actual financial cost of marking a particular parameter from blood, which may vary between 5PLN up to over 50PLN, and also on a blood sample analysis time required for marking that parameter, which may vary from less than an hour up to a week. As the condition is chronic it may be thought that time is not a critical parameter, but if it would be possible to have a diagnosis within an hour or two it would be possible to diagnose a patient on an outpatient basis and would also have a positive psychological meaning for a patient (not having to wait for a days for a diagnosis to come).

Table 1 presents number of examined patients for each fibrosis stage (F0..F4) and Table 2 presents characteristics of acquired medical records and adopted cost of marking each of the parameters.

Table 1 Number of patients with given fibrosis stage [n (%)]

<i>F0</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>
2 (2%)	34 (33%)	5 (5%)	16 (15%)	46 (45%)

Table 2 Blood test results characteristics [name / mean (std. deviation) / associated cost].

HB (g/l)	14 (1.80)	1
RBC ($10^6/\mu\text{l}$)	4 (0.69)	1
WBC ($10^3/\mu\text{l}$)	6 (2.35)	1
PLT ($10^3/\mu\text{l}$)	161 (73.08)	1
PT (sec.)	13 (9.99)	6
PTP (%)	91 (17.23)	6
APTT (sec.)	37 (7.29)	6
INR	1 (0.16)	6
ASPT (IU/l)	69 (55.14)	30
ALAT (IU/l)	77 (65.43)	30
ALP (IU/l)	105 (57.70)	5
BIL (mg/dl)	2 (2.42)	6
GGTP (IU/l)	94 (101.31)	5
KREA (mg/dl)	1 (0.24)	4
GLU (mg/dl)	93 (17.44)	4
Na (mmol/l)	138 (3.28)	5
K (mmol/l)	5 (5.74)	5
Fe (mmol/l)	92 (63.51)	5
CRP (IU/l)	5 (28.77)	11
TG (mg/dl)	107 (53.76)	5
CHO (mg/dl)	191 (53.23)	5
Ur. acid (mg/dl)	6 (1.35)	4
TP (g/dl)	7 (0.84)	4
TIBC	316 (95.02)	5
Neutr ($10^3/\mu\text{l}$)	3.42 (1.35)	25
Lymph ($10^3/\mu\text{l}$)	2.05 (0.55)	1
Mono ($10^3/\mu\text{l}$)	0.58 (0.19)	1
Eos ($10^3/\mu\text{l}$)	0.17 (0.13)	1
Baso ($10^3/\mu\text{l}$)	0.03 (0.02)	1
Albu (%)	59.3 (7.04)	5
Glb. $\hat{\text{I}}\acute{\text{s}}1$ (%)	2.9 (1.31)	30
Glb. $\hat{\text{I}}\acute{\text{s}}2$ (%)	8.8 (2.57)	30
Glb. $\hat{\text{I}}\acute{\text{s}}$ (%)	10.8 (1.63)	30
Glb. $\hat{\text{I}}\acute{\text{s}}$ (%)	18.4 (6.89)	30

3 Cost Sensitive Hierarchical Classifier

The idea behind the multi-stage approach [16] resolves around breaking up a complex decision into several simpler tasks [4]. The decision tree classifiers are the most popular algorithms devoted to the multi-stage pattern recognition [13]. The synthesis of such a hierarchical classifier is a complex problem [12]. It involves specification of the following components:

- design of a decision tree structure,
- feature selection used at each nonterminal node of decision tree,
- choice of decision rules for performing the classification.

The central choice designing a decision tree is selecting the best attribute to test at each node of such a hierarchical classifier [19]. In this paper we use the information gain that measures how well the given attribute separates the training examples according to the target classification. This measure is based on the Shanon's entropy of set S :

$$\text{Entropy}(S) = \sum_{i=0}^M -p_i \log_2 p_i, \quad (1)$$

where p_i is the proportion of S belonging to class i .

The Information Gain of an attribute A relative to the collection of examples S , is defined as

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{c \in \text{values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v), \quad (2)$$

where $\text{values}(A)$ is the set of all possible values for attribute A and S_v is the subset of S for which $A = v$.

Let us now introduce a cost-sensitive modification of such a hierarchical classifier. The motivation behind such a choice of the classification method lies in the fact, that we would like to make our proposed non-invasive examination as affordable for the patient as possible. Therefore we would like to construct a classifier exhibiting a high possible accuracy that uses features with low total cost [18].

We use the the Núñez's proposition, known as EG2 [9] that can be formulated as follows:

$$\text{ICF}(S, A) = \frac{2^{\text{Gain}(S, A)}}{(\text{Cost}(A) - 1)^\omega} \quad (3)$$

where ω is the strength of the bias toward the lowest cost attributes; in case of $\omega = 0$ the feature acquisition cost is ignored and ICF as the same features as Gain function, if $\omega = 1$ - mentioned cost plays the most important role.

We propose to use different values instead of information gain function to determine which attribute has to be chosen for the node creation [11]. Firstly, we add the following input parameters:

- TOTAL-COST, which means cost connected with classifier exploitation (as a general initial value is fixed by expert),
- MAX-COST-LIMIT which means maximum cost limit connected with classifier exploitation,
- STEP , which is the increase of the TOTAL-COST in the next induction step,
- EXPECTED-QUALITY, which denotes the expected performance of the inducted cost sensitive decision tree (that can be expressed as accuracy, error, sensitivity or any other measure).

In our approach we propose to select a subset of attributes that total cost sum does not exceed the cost limit given as a parameter for the classifier. In case when set of attributes is empty or cost of the previous chosen attributes and cost of any remaining one exceed the cost limit we return the single node tree with label equal to the most common value of labels in all examples.

We add a procedure to evaluate an obtained cost sensitive decision tree. If the overall quality of the classifier inducted at the n -th step is lower than the expected quality we add a $n + 1$ step by adding another node and increase the total cost by the cost associated with the newly selected attribute. This is done as long as we achieve the expected quality or the total cost exceeds the maximum cost limit. In case when the total cost is greater than the limit and the quality is not satisfactory we return the tree from the last step and inform the end-user that the inducted cost sensitive tree does not fulfill the given assumptions.

4 Experimental Investigations

4.1 Set-Up

The aim of the experiment was to compare errors and size of the decision tree classifiers obtained via C4.5 procedure which consider the classification (attribute acquisition) cost limit. We carried out two group of experiments.

The conditions of the experiment were as follow:

- all experiments were made for different limits of cost and ω values for EG2 modification.
- for experiments we chose pruned decision tree obtained via C4.5 [14].
- for pruning the Rule Post Pruning method was used.
- all experiments were carried out using modified Quinlan's implementation of C4.5 and own software created in the R environment [17]. For this purpose we modified the C4.5 algorithm source code.
- Probabilities of errors of the classifiers were estimated using the 10-fold cross-validation method.

4.2 Results

The correlation between the given maximum cost limit and achieved accuracy is presented in Fig. 1, while Fig. 2 shows how the size of the tree (i.e., how many features are used) depends on the given cost limit.

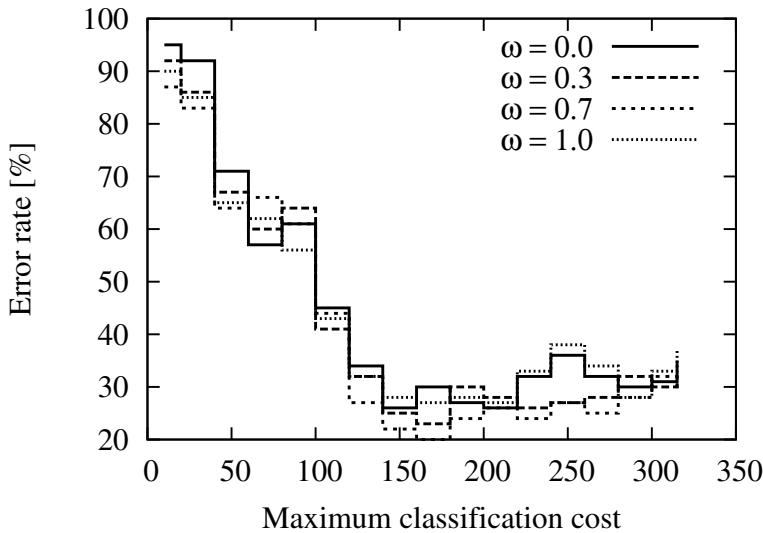


Fig. 1 Classification error with the respect to the maximum classification cost for the tested dataset

4.3 Results Discussion

From the Fig. 1 we can see that it is possible to build an accurate classifier while reducing the overall cost of the examination. The overall cost of all of features equals to 315. By the usage of our proposed cost-sensitive hierarchical classifier we were able to find a trade-off between the accuracy and cost. The optimal solution returns a classification accuracy of circa 80%, while having a total cost of used features equal to 160. This allows for almost a half of the reduction of the total cost of the examination.

Of course it is possible to have a lower cost if we input a stricter cost limit to the classifier. It is possible to further reduce the cost to up to 100, while maintaining the accuracy on the level about 70%. Further decrease of the cost is connected with the significant drop of accuracy and therefore is of no practical clinical value and cannot be implemented in the proposed decision support system.

Let us now take a look on the size of built decision trees. For optimal solution classifier uses circa 9 features (this is an average as the size of the tree may differ

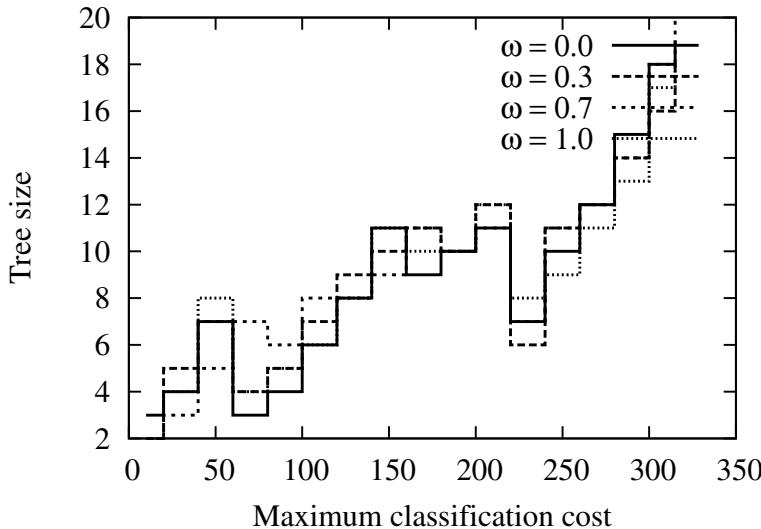


Fig. 2 Decision tree size with the respect to the maximum classification cost for the tested dataset

for each of the validation folds). The lowest maximum cost that still provides a satisfactory accuracy is based on 6 features. Interestingly for higher cost limit the number of selected features drop - this indicates that in such a case the classifier have selected few very costly features - for low values of upper bound on the total cost such situation must be omitted. Therefore classifier selects more weaker features of low cost, which combination still provides a high discriminative power.

The influence of cost parameter ω on the performance of our classifiers is not as big as we expected. When discarding the influence of the cost the tree tends to consist of few expensive features. Introducing cost factor prevents this selection - yet there are no significant differences between the tested values of ω .

Finally we would like to identify the most frequently used features. It is obvious that for each fold of the cross validation the selected subset may differ but following features were selected in a stable way (eight out of ten times or more): ALP, APTT, ALAT, GGTP, Ur. Acid, TIBC, Glb α_1 , Glb α_2 .

5 Conclusions

By our experiments, we have shown that, what is obvious, the classification error rate decreases with the growth of the cost limit, but there is a point from which error rate not only stops to decrease but starts to grow. The optimal cost is slightly above the 50% of the cost of marking all the parameters and the decision tree size

for this optimal cost is 9, so the optimal classifier uses only 9 out of 34 parameters to evaluate the result. In about 80% of validation folds 8 out of 9 chosen parameters were the same, so it is possible to define a stable, minimal subset of parameters which is required for the purpose of evaluation the stage of Liver Fibrosis.

For the further research it is possibly worth to include into the Cost Factor also the dependencies between blood test parameters as some requires other to be made before and some are always made together from one sample. Also volume of blood sample required for examination could be considered as a cost affecting parameter.

Acknowledgements. Bartosz Krawczyk and Michał Woźniak are supported by the Polish National Science Center under a grant N N519 650440 for the period 2011-2014.

References

1. Bedossa, P., Dargre, D., Paradis, V.: Sampling variability of liver fibrosis in chronic hepatitis. *Hepatology* 38, 1449–1457 (2003)
2. Bedossa, P., Poynard, T.: An algorithm for the grading of activity in chronic hepatitis. *The Metavir Cooperative Study Group* 24, 289–293 (1996)
3. BioPredictive. Website, <http://www.biopredictive.com/intl/physician/fibrotest-for-hcv/>
4. Burduk, R., Zmysłony, M.: Decomposition of classification task with selection of classifiers on the medical diagnosis example. In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B. (eds.) HAIS 2012, Part II. LNCS, vol. 7209, pp. 569–577. Springer, Heidelberg (2012)
5. Castéra, L.: Non invasive assessment of liver fibrosis in chronic hepatitis. *Liver International* 5(2), 625–634 (2002)
6. Siemens Healthcare. Website, <http://healthcare.siemens.com/clinical-specialities/liver-disease>
7. iQu Ltd. Website, <http://www.iqur.com/ELFTest.html>
8. Krawczyk, B., Woźniak, M., Orczyk, T., Porwik, P., Musialik, J., Blońska-Fajrowska, B.: Classification techniques for non-invasive recognition of liver fibrosis stage. *Journal of Medical Informatics & Technologies* 20, 121–127 (2012)
9. Nuenz, M.: The use of background knowledge in decision tree induction. *Machine Learning* 6, 231–250 (1991)
10. Orczyk, T., Pałys, M., Porwik, P., Musialik, J., Blońska-Fajrowska, B.: Simple and non-invasive liver fibrosis stage prediction method. *Journal of Medical Informatics & Technologies* 17, 227–232 (2011)
11. Penar, W., Woźniak, M.: Cost-sensitive methods of constructing hierarchical classifiers. *Expert Systems* 27(3), 146–155 (2010)
12. Podolak, I.T., Roman, A.: Risk function estimation for subproblems in a hierarchical classifier. *Pattern Recognition Letters* 32(15), 2136–2142 (2011)
13. Quinlan, J.R.: Induction of decision trees. *Machine Learning* 1(1), 81–106 (1986)
14. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Series in Machine Learning. Morgan Kaufmann Publishers (1993)

15. Regev, A., Berho, M., Jeffers, L.J., Milikowski, C., Molina, E.G., Pyrsopoulos, N.T., Feng, Z.Z., Reddy, K.R., Schiff, E.R.: Sampling error and intraobserver variation in liver biopsy in patients with chronic hcv infection. *The American Journal of Gastroenterology* 97(10), 2614–2618 (2002)
16. Senator, T.E.: Multi-stage classification. In: Proceedings - IEEE International Conference on Data Mining, ICDM, pp. 386–393 (2005)
17. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2008)
18. Ting, K.M.: An instance-weighting method to induce cost-sensitive trees. *IEEE Transactions on Knowledge and Data Engineering* 14(3), 659–665 (2002)
19. Weiss, Y., Elovici, Y., Rokach, L.: The cash algorithm-cost-sensitive attribute selection using histograms. *Information Sciences* 222, 247–268 (2013)

Part VI

Miscellaneous Applications

A Blinking Measurement Method for Driver Drowsiness Detection

Belhassen Akrout and Walid Mahdi

Abstract. The increasing number of accidents is attributed to several factors, among which is the lack of concentration caused by fatigue. In This paper, we describe the approach developed to detect the driver's drowsiness state from a video-based system to alert him and also reduce the number of accidents. Our approach uses a noninvasive method which excludes any human related elements. The latter calculates geometric descriptors. We analyze the signal extracted from the previous step by combining the two methods EMD (Empirical Mode Decomposition) and BP (Band Power). This analysis is confirmed by the SVM (Support Vector Machine) to classify the state of alertness of the driver.

1 Introduction

Road safety is an issue that has been raised by several countries to increase the number of accidents. In this paper, we focus on the problem of detecting driver drowsiness state to reduce the number of accidents and improve road safety. Many efforts have been made to detect the state of driver's drowsiness. Some methods are based on EEG signal analysis [14] [15] and others analyzes the video driver [22] [23] [1] [3] [4] [24] [25]. We are interested in our work to examine the measures related to the speed of eye closure from a sequence of images to determine the driver's state.

Approaches to drowsiness can be divided into two categories. The first category, called mono-variables, calculates the time of eye closure. Indeed, the size of the iris surface depends on its state in the video. This analysis can be used to determine the

Belhassen Akrout · Walid Mahdi
Laboratory MIRACL, Institute of Computer Science and Multimedia of Sfax,
Sfax University, Tunisia
e-mail: akroud_belhassen@yahoo.fr, walid.mahdi@isimsf.rnu.tn

state of the eye (open or close) [1] [2] [3]. Other studies calculate the distance between the upper and lower eyelids to locate eye blinks. This distance decreases if the eyes are closed and increases when they are open [4] [5] [6] [7]. These approaches can alert the driver in case of prolonged eye closure. The duration of eye closure used as an indication varies from one work to another. Sarbjit [2] considers that a person is asleep if the eyes remain closed for a period of 5 to 6 seconds; in this case, we speak of total sleep. On the other side, the micro-sleep is detected if the driver goes through a state of sleep for a short time (2-3 seconds). For Horng [1], the driver is considered drowsy if he closes his eyes for 5 consecutive frames. Hongbiao [6] considers that the state of drowsiness is determined when the distance between the eyelids is less than 60% for a period of 6.66 seconds. Yong [7] divides the opening state of the eyes into three categories (open, half open, closed). This division allows concluding the driver's drowsiness if the eyes are kept closed more than four consecutive images or eyes move from a state of half open to a closed state for eight successive images. Besides, the percentages of detection of fatigue vary in literature. Yong reached 91.16% of correct average rate for recognition of the condition of the eyes. As for Horng, he explains that the average accuracy rate for detection of fatigue can reach 88.9%. Wenhui [8] reached 100% of rate of detection efficiency. All these works compute their results with subjects that their numbers varie from two to ten people (four individuals for Wen-Bing [1], five subjects Hiroshi [1] for ten Tnkehiro [4] and only two for Yong [7]).

The second type of approach is called multi-variable. In this context, the maximum speed reached by the eyelid when the eye is closed (called velocity) and the amplitude of blinking calculated from the beginning of blink until the maximum blinking are two indications that have been studied by Murray [9]. The latter shows that the velocity amplitude ratio (A/PCV) is used to prevent the driver one minute in advance. Takuhiko [10], uses an infrared camera and suggests five levels of vigilance namely non-drowsy, slightly drowsy, sleepy, rather sleepy, very sleepy and asleep. Takuhiko proposes several criteria: The percentage of eye opening, the number of blinking, the duration of eye closure, the blinking interval and the surface of eye opening. This work has been tested using a base of ten subjects. A correlation rate that is greater than 0.90 is achieved. Picot presents a synthesis of different sizes as the duration to 50%, the PERCLOS 80%, the frequency of blinking and the velocity amplitude ratio. Picot [11] shows that these criteria are more relevant to the detection of drowsiness. These variables are calculated every second on a sliding window of the length of 20 seconds. They are fused by fuzzy logic to improve the reliability of the decision. This study shows a percentage of 80% of good detections and 22% of false alarms.

In the case of multivariate approaches, we find that some methods are based on the analysis of the EEG signal. These invasive methods require a technical cooperation between the hardware and the driver. They use a large number of parameters, which require more data for learning. Nevertheless, video-based approaches, rest on the segmentation of the iris of the eye so as to extract the features for the subsequent

steps. The segmentation of the iris is calculated by the difference between 2 images by using an infrared camera. The disadvantage of this method lies in the noise sensitivity of the luminance. In this context, we support the localization of the iris, the method of Circular Hough Transform. This method shows sturdiness in the face of the desired shape, an ability to adapt even to images with poor or noisy quality as well as an identification of all directions thanks to the use of a polar description. Single-variables Approaches, present a very high rate of successes detection, but the disadvantages is that the detection of the state of drowsiness is determined in a very advanced stage. We also concentrate in our work to study the characteristics that have the goal to predict the driver state before sleeping, and analyzing the speed of closing eyes. It is from these remarks that we develop in this paper our approach called spatiotemporal drowsiness detection by calculating descriptors from a sequence of images in a specific time.

2 Proposed Approach

This paper presents an approach for detecting drowsiness of a driver by studying the behavior of conductor eyes in real time by an RGB camera (Fig. 1). This approach requires a critical step presumed through the automatic face detection, first, and the detection of the box that encompasses both eyes, in the step that follows.

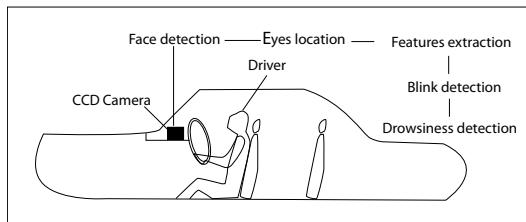


Fig. 1 Detection scheme of drowsiness

2.1 Face Detection and Eyes Localization

In order to come to delight the face and the eyes, our approach exploits the object detector of Viola and Jones that is about a learning technique based on Haar features. This method [12] uses three concepts: the rapid extraction of features using an integral image, a classifier based on Adaboost and the implementation of a cascade structure. Fig.2 shows the detection result of the face and the eyes.



Fig. 2 Detection results of both face detection and eye localization

2.1.1 Iris and Both Eyelids Detection

With reference to the observation of the eye, we note that human eyes are characterized by horizontal contours representing the eyelids and the wrinkles or vertical contours as the ones of the iris. The application of two-scale Haar wavelet allows extracting the vertical, horizontal and diagonal contours. The vertical contours are used in the localization of the iris following application of the Circular Hough Transform. The use of the wavelet allows us to highlight the contours that we want to spot frequently. In our case, the scale of the second rate improves the contours of the iris and the two lids which are going to be detected.

Edge Extraction Based on 2D Haar Wavelet

The Haar wavelet [20] allows us to split the image to find the vertical and horizontal details for the detection of the iris and both eyelids. The wavelet transform is characterized by its multi-resolution analysis. It is a very effective tool for noise reduction in digital image. We can also ignore certain contours and keep only the most representative ones. This type of analysis is allowed by the multi-resolution.

Iris and the Two Eyelids Detection Based on Circular Hough Transform

In general, the Circular Hough Transform [13] has two spaces, the space XY and parameter space which varied according to the detected object. Our approach involves the detection of the iris by applying the Hough transform on the vertical details of the eye. Both eyelids are located using the Circular Hough Transform on the image of horizontal details of the Haar wavelet decomposition (Fig.3).

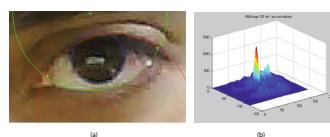


Fig. 3 Detection of the iris and the two eyelids. (a) Result of contour detection by the Circular Hough Transform, (b) 3D shape of the accumulator for the detection of the iris.

2.2 Geometric Features Extraction

With reference to the detection of the iris, the upper eyelid and lower one, we can extract geometric features able to characterize the state of drowsiness of a driver. We propose two geometric features D_{ih} and D_{ib} (Fig.4).

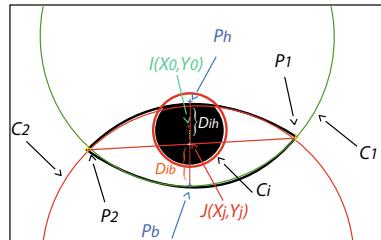


Fig. 4 Representation of features from the figure of the eye

These features represent the distance between the point J and respectively the point P_h and P_b , Eq.(1).

$$D_{ih} = \sqrt{(x_j - x_1)^2 + (y_j - y_1)^2} \text{ and } D_{ib} = \sqrt{(x_j - x_2)^2 + (y_j - y_2)^2} \quad (1)$$

We note that, P_h and P_b are calculated from the intersection between the straight line which passes through J and perpendicular to the line (P_1P_2) and the two circles C_1 and C_2 , where the equation of one of the two circles is follow as:

$$R^2 = (x_0 - x)^2 + (y_0 - y)^2 \quad (2)$$

And the perpendicular to line (P_1P_2) is calculated by applying the by Eq.(3).

$$y = ax + b \Rightarrow y^2 = ax^2 + b^2 + 2axb \quad (3)$$

By replacing y^2 into R^2 after development we obtain the Eq.(4).

$$(1+a)x^2 + (2ab - 2x_0 - 2y_0)x + (b^2 + x_0^2 + y_0^2 - R^2) = 0 \quad (4)$$

The straight line (P_1P_2) is determined after the detection of two points P_1 and P_2 (Fig.4) which represent the intersection of two circles C_1 and C_2 such as the squared equation of circle C_1 is:

$$R_1^2 = (x_1 - x)^2 + (y_1 - y)^2 \quad (5)$$

And the squared equation of circle C_2 is written as follow:

$$R_2^2 = (x_2 - x)^2 + (y_2 - y)^2 \quad (6)$$

The difference between R_1^2 and R_2^2 allows to determinate the tow points P_1 and P_2 by the Eq.(7).

$$x^2 A + x + [x_1^2 + y_1^2 + N^2 - R_1^2 - 2y_1N] - R_2^2 + R_1^2 = 0 \quad (7)$$

Such us:

$$N = \frac{R_2^2 - R_1^2 - x_2^2 + x_1^2 - y_2^2 + y_1^2}{2(y_1 - y_2)} \quad (8)$$

While:

$$A = \left[\frac{(x_1 - x_2)^2}{(y_1 - y_2)^2} + 1 \right] \quad (9)$$

And:

$$B = \left[\frac{2y_1(x_1 - x_2)}{(y_1 - y_2)} - \frac{2N(x_1 - x_2)}{(y_1 - y_2)} - 2x_1 \right] \quad (10)$$

2.3 Experiments and Prototypes

In this section, we describe the experimental studies we conducted to validate the two features D_{ih} and D_{ib} previously proposed. In order to produce realistic data, a human subject is placed in front of our system to simulate different possible movements of the head, the eyelids and the positions of the iris, probably related to different states of fatigue. This experiment consists of studying the temporal variation of both features and normalizing the initial state of the Eq.(11).

$$f(x_t) = (D_{ih}^t + D_{ib}^t)/V_i \quad (11)$$

The initial value V_i is calculated at the beginning of the algorithm when the eyes were open nearly 75% (Fig.6).



Fig. 5 Different states of an open eye

The nature of the signal (Fig.6) obtained is non-stationary and non-linear.

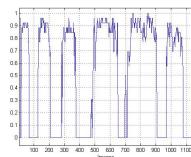


Fig. 6 An example of $f(x_t)$ calculated states of an open eye

The Fourier Transform or Wavelets, made the time-frequency analysis of signals possible. These latter methods do not properly handle non-stationary and non-linear signals. It is from these arguments that we justify the adoption of a method called Empirical Mode Decomposition [16] (EMD). This method does not depend on a core of basic function (such as Wavelets) and depend on the original signal. EMD method decomposes the $f(x_t)$ signal into a local average $a(x_t)$, which is the average of the upper and lower envelopes of $f(x_t)$ and another element oscillating intensely $IMF_i(x_t)$. With iteration we can repeat the above procedure until a non-oscillating residue $r(x_t)$ Eq.(12).

$$f(x_t) = \sum_i IMF_i(x_t) + r(x_t) \quad (12)$$

The EMD is based on the extraction of the function in which all maxima are positive and all minima are negative (Intrinsic Mode Function) and the sieving step (Sifting Process) which represents the subtraction of the signal $f(x_t)$ from the local average $a(x_t)$ so as to write the original signal as combinations finished oscillation Eq. (12). Interpolation method makes sometimes the program not converging. This issue is addressed through improved this method by increasing the accuracy of calculation [19].

The decomposition of the signal (Fig.7) shows a strong oscillation in the first $IMF_1(x_t)$ when the driver is drowsy. This observation argues for the choice of the first $IMF_i(x_t)$ to extract the most relevant information of the signal by combining the EMD and BP algorithm on a sliding window of a second, since a normal blink lasts no more than 0.2 seconds [21].

The BP method [17] applies a band pass filter passing only a frequency interval, which is the band of the cut between the high and low frequency, this interval, is averaged after it is squared. The descriptors extracted are validated by the results of classification with the SVM algorithm [18], in order to determine the status of driver drowsiness. Fig.8 shows the states of good detections drowsiness presented by the value 1 and vigilance by the value of 0.

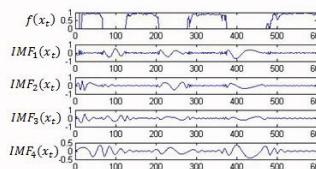


Fig. 7 Decomposition of the signal $f(x_t)$ to $IMF_i(x_t)$

A few false alarms appear in the result of the classification. These false alarms (Image 2196 to image 2220 of Fig. 8) are due to a blurring of the subject in the video recorded. The feature extraction method with BP and EMD, and the classification performed by the SVM algorithm shows satisfactory results. This new approach validates the relevance of the two geometric descriptors.

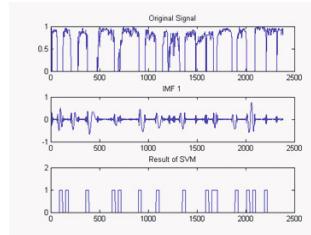


Fig. 8 Result of fatigue detection by SVM classification

Table 1 Result of Fatigue Detection

Videos	Real Drowsiness	Generated alarm	Correct alarm	Recall	Accuracy
1	10	9	9	0.9	1
2	13	14	12	0.92	0.85
3	7	7	6	0.85	0.85
4	12	14	11	0.91	0.78
5	9	8	8	0.88	1
6	7	8	6	0.85	0.75
7	6	6	6	1	1
8	8	8	7	0.87	0.87
Average				0.897	0.887

Table 1 shows the detection result of fatigue on the eight test videos. The expert advice in this step is necessary to determine the actual drowsiness of a driver. The recall rate is calculated from the correct alarm index and those determined by the specialist. The accuracy is calculated as a ratio between the correct and generated alarms.

The results of the proposed approach, for detecting the driver drowsiness state, have a good average Recall and Precision (near 0.9). However, the errors generated alarms are explained by the effect of blurring recorded with RGB camera. These errors influence the detection of the iris and then the values of the descriptors.

3 Conclusion

This paper presents an approach to detection of reduced alertness, based on computer vision. Our system uses the state of the eyes by the video analysis of several topics. The steps of drowsiness detection, locate face and eyes driver using Haar features in the first. The circular Hough Transform allows the detection the iris center and the points of intersection for the two lids (Fig. 7), in order to operate the

two geometric descriptors. The result of signal after the extraction of two geometric descriptors is non-linear and non-stationary. A BP and EMD method allows the analysis and extraction of this type of signal. This analysis is confirmed by the SVM method to find satisfactory results for the benefit of classification. The results of this approach show a good accuracy rate (88.7%). To improve them we need to determine the yawn states, monitoring and estimating the 3D pose of the face that can be important clues about the state of driver alertness. All these methods present the goal of future works to improve our results

References

1. Horng, W., Chen, C., Chang, Y., et al.: Software engineering – Driver fatigue detection based on eye tracking and dynamic template matching. In: IEEE International Conference on Networking, Sensing and Control. IEEE, New York (2004)
2. Sarbjit, S., Nikolaos, P.: Monitoring Driver Fatigue Using Facial Analysis Techniques. Intelligent Transportation Systems. ITS, Japan (1999)
3. Hiroshi, U., Masayuki, K.: Software engineering – Development of drowsiness detection system. Vehicle navigation and information systems conference. VNISC, Japan (1994)
4. Tnkehiro, I., Shinji, M., Kduo, K., Tomoaki, N., Shin, Y.: Driver Blink Measurement by the Motion Picture Processing and its Application to Drowsiness Detection. In: IEEE International Conference on Intelligent Transportation Systems. IEEE, Singapore (2002)
5. Masayuki, K., Hideo, O., Tsutomu, N.: Adaptability to ambient light changes for drowsy driving detection using image processing. UC Berkeley Transportation Library (1999)
6. Hongbiao, M., Zehong, Y., Yixu, S., Peifa, J.: A Fast Method for Monitoring Driver Fatigue Using Monocular Camera. In: Proceedings of the 11th Joint Conference on Information Sciences. JCIS, China (2008)
7. Yong, D., Peijun, M., Xiaohong, S., Yingjun, Z.: Driver Fatigue Detection based on Eye State Analysis. In: Proceedings of the 11th Joint Conference on Information Sciences. JCIS, China (2008)
8. Wenhui, D., Xiuojuan, W.: Driver Fatigue Detection Based On The Distance Of Eyelid. In: IEEE Workshop Vlsi Design and Video Tech. IEEE, China (2005)
9. Murray, J., Andrew, T., Robert, C.: A new method for monitoring the drowsiness of drivers. In: International Conference on Fatigue Management in Transportation Operations. CFMTO, USA (2005)
10. Takuhiro, O., Fumiya, N., Takashi, K.: Driver drowsiness detection focused on eyelid behavior. In: 34th Congress on Science and Technology of Thailand. CSTT, Thailand (2008)
11. Picot, A., Caplier, A., Charbonnier, S.: omparison between EOG and high frame rate camera for drowsiness detection. In: IEEE Workshop on Applications of Computer Vision. IEEE, USA (2009)
12. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. of CVPR (2001)
13. Cauchie, J., Fiolet, V., Villers, D.: Optimization of an Hough transform algorithm for the search of a center. Pattern Recognition (2008)
14. Latif, M., Sanei, S., Chambers, J.: Localization of abnormal EEG sources incorporating constrained BSS. In: International Conference on Artificial Neural Networks (2005)

15. Larue, G.S., Andry, R., Anthony, P.: Driving performance impairments due to hypovigilance on monotonous roads. *Accident Analysis and Prevention* (2011)
16. Huang, N.E., Shen, Z., Long, S.R., Wu, M.L., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of Royal Society of London. London* (1998)
17. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. *Proceedings of the IEEE* (2001)
18. Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J.: *Least Squares Support Vector Machines*. World Scientific, Singapore (2002)
19. Boustane, T., Quellec, G., Chainais, P.: Implantation de la methode EMD en C avec interface Matlab. Project report ISIMA, France (2004)
20. Porwik, P., Lisowska, A.: *The Haar Wavelet Transform in Digital Image Processing: Its Status and Achievements. Machine GRAPHICS and VISION*. MGV, Poland (2004)
21. Kojima, N., Kozuka, K., Nakano, T., Yamamoto, S.: Detection of Consciousness Degradation and Concentration of a Driver for Friendly Information Service. In: *Vehicle Electronics Conference Proceedings of the IEEE International*. IEEE, Japan (2001)
22. Garcia, I., Bronte, S., Bergasa, L.M., Almazan, J., Yebes, J.: Vision-based drowsiness detector for real driving conditions. In: *IEEE Intelligent Vehicles Symposium*. IEEE, Spain (2012)
23. Devi, M.S., Choudhari, M.V., Bajaj, P.: Driver Drowsiness Detection Using Skin Color Algorithm and Circular Hough Transform. In: *Fourth International Conference on Emerging Trends in Engineering and Technology*. CETET, Mauritius (2011)
24. Akroud, B., Mahdi, W.: Drowsiness Detection Based on Video analysis Approach. In: *The 8th International Conference on Computer Vision Theory and Applications*. VISAPP, Barcelona, Spain (2013)
25. Akroud, B., Mahdi, W.: Vision based approach for driver drowsiness detection based on 3D head orientation. In: *The 7th FTRA International Conference on Multimedia and Ubiquitous Engineering*. MUE, Seoul, Korea (2013)

Description of Human Activity Using Behavioral Primitives

Piotr Augustyniak

Abstract. Human activity is a subject of tracking and recognition in various aspects including: public security, health lifestyle or home monitoring of elderly. A multi-modal surveillance system is proposed to recognize the action and to assess the similarity of temporal behavioral patterns. The system uses sensor networks, automatic measurement module and decision making procedure to recognize the potentially dangerous events. It uses behavioral primitives (as positions, movements or vital signs) and their temporal relations to determine the current activity of the subject.

The idea of decomposition of human behavior description is developed throughout this paper. In principles it originates from the signal theory and assumes that any behavioral pattern can be represented by a linear combination of independent elementary actions. These actions should be carefully selected to provide a minimal redundancy and ease the measurement and robust recognition in real systems.

1 Introduction

Telemedical tools for distant monitoring and management of outpatients became practical during last years in particular for diabetic, asthmatic or cardiac patients [6], [17], [5], [16]. Combination of medical methods and traditional (video)surveillance technologies led to extending of the monitoring range beyond the purely medical purposes. It can be applied in service of virtually any individual for health-based supervision of the body and behavior, and targeted to a wide range of customers in the ageing society such as families or independently living people.

Piotr Augustyniak

AGH University of Science and Technology, 30 Mickiewicza Ave. 30-059 Kraków Poland
e-mail: august@agh.edu.pl

1.1 Combining Telemedical and Surveillance Techniques

While the monitoring of asthmatic or diabetes uses conventional event-based connections, cardiovascular diseases require a more sophisticated approach [4], [8], [1], where high data volume or patient-side interpretation must be supported for seamless monitoring and instant alerting. Because of the cardiovascular representation of motion, feelings or sleep, cardiac monitoring sensors may also be useful for behavior tracking [2]. Most systems being in use today are dedicated to a particular purpose and use a closed architecture, however a need for adaptive sensor sets and support of an open specification-based multimodal record has been clearly addressed in recent papers.

Surveillance or objects tracking systems are wide spread in supermarkets, airports, delivery and manufacturing automation. Various technologies are used for object identification including video analysis (e.g. code bars), approach detection (e.g. radio-frequency identification tags) or interactions with static elements of the environment. The image-based methods assume the invariance of visual features of objects and despite their uneven performance in detecting a human in unstable natural lighting conditions, are today the most common solution for supervision of people [13], [9]. In many presence detection systems, visual sensors are complemented with radio wave or ultrasound-based tactless detectors and with sensitive elements of the subject's environment.

1.2 Multimodal Representation of Behavior

Combination of these well mastered techniques leads to a new generation of human behavior surveillance systems expected to detect any danger events in an everyday life and to prevent its consequences [7], [11]. The system consists of an adaptive smart sensor array (consisting of wearable and infrastructure embedded sensors), behavior analysis subsystem detecting and classifying any outstanding event and a habits synthesis subsystem gathering and generalizing information about subject-specific features. The behavioral record plays a central role in the whole system, thus the proper representation of the human activity is a key issue to be solved.

The information gathered by the sensor array is instantly processed to identify the subject status. The status includes selected diagnostic parameters and information about subject's motion and interaction with the elements of house infrastructure. The behavioral pattern is a multimodal timeline record of subject's status including: probability-ordered list of current status, average and standard deviation values of expected duration, probability-ordered list of subsequent status and optional pointers to detailed subject-related sensor data (video sequence, sounds, acceleration, positioning or cardiac signals). The redundancy of data acquired from different sensing modes are further used by the analytical subsystem to:

- maintain the continuity of monitoring in case selected data are not available (e.g. the subject is outdoor), and
- identify multiple subjects supervised in a common area.

2 The Idea of Decomposition of Behavioral Description

The human is performing very complicated actions in his or here everyday life, making their machine recognition and classification very difficult. Moreover, the action performed is influenced by subject's specificity (e.g. habits, conditions, health state), and by a random human factor, thus the realizations differ even if repeated in identical laboratory conditions. The machine representation of human action is further influenced by limitations of measurement equipment or procedures. Consequently, behavior classification algorithms cannot be directly based on classification of images, sounds or other signals.

To overcome these issues a three step algorithm was proposed:

1. the gathered data is processed accordingly to the measured modality (that's why we call the sensors 'intelligent'),
2. the multimodal data are combined to recognize the behavioral primitives, and
3. the compound human action is integrated accordingly the contribution of each primitive.

Before the algorithm can be put into practice, an universal set of behavioral primitives should be defined. From the signal theory point of view, the primitives need to be independent (i.e. uncorrelated) and constitute a complete set (i.e. be able to represent any possible behavior). For a practical implementation, the definition should respect the limitations of measurement and processing in the sensor network.

2.1 Selecting a Decomposition Base for the Behavioral Record

The decomposition base for a specific behavioral record is a set of primitives such that any behavior b belonging to the space of expected behaviors B could be represented as integral of these primitives:

$$\forall_{b \in \mathbf{B}} : b = \int p(k) \cdot n(k) dk \quad (1)$$

In practical implementations, due to limited dimensionality of decomposition space, the integral should be replaced by linear combination.

$$\forall_{b \in \mathbf{B}} : b = p_0 \cdot n_0 + p_1 \cdot n_1 + \dots + p_k \cdot n_k \quad (2)$$

where k is the space dimensionality, $p_k \in \mathbf{P}$ are decomposition base, and $n_k \in \mathbf{N}$ are decomposition coefficients.

The example set of poses proposed as behavioral primitives of human activity decomposition base is presented in Figure 1. Although a real human action is rarely simple enough to be represented by a single primitive, our results show, that in most cases a robust representation can be achieved with very limited number of primitives.

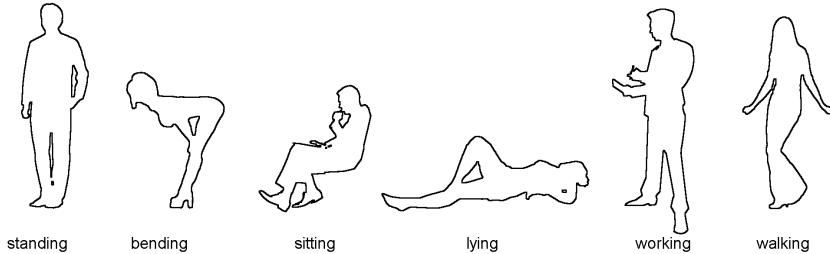


Fig. 1 Set of poses proposed as behavioral primitives of human activity decomposition base

The necessity of completeness of the decomposition space assumes that all elements of subject's behavior are predictable and thus is hardly feasible in practice. To make the practical decomposition complying to the theory, all unexplained events are classified to a complementary category 'others' (for convenience indicated by index 0 in each dimension).

$$\forall b \in \mathbf{B} : b = \sum_{k=1}^K p_k \cdot n_k + p_0 \cdot n_0 \quad (3)$$

where K is the number of primitives in decomposition base, p_0 stands for all unexplained actions, and n_0 is the unexplained contribution in analyzed action.

2.2 Basic Operations on a Decomposed Behavioral Record

This approach led us to formulate a Fourier-like discrete multidimensional decomposition of human state description. Each dimension corresponds to a measurement modality (e.g. position, acceleration, heart rate) and $S()$ being the current state of the subject is expressed by coordinates in a decomposition space:

$$S(m_1 \pm \delta m_1, m_2 \pm \delta m_2, \dots, m_M \pm \delta m_M) = \quad (4)$$

where M is the number of measurement modes, m_k is the most prominent primitive in k -th mode, and δm_k is the uncertainty (i.e. contribution of other significant primitives) in k -th mode.

Assuming orthogonality of the modality space, the distance D between two states S_1 and S_2 of the same subject in selected moments, or between the current and the reference states may be expressed by a conventional Euclidean formulae:

$$D(S_1, S_2) = \sqrt{(m_{1,2} - m_{1,1})^2 + (m_{2,2} - m_{2,1})^2 + \dots + (m_{M,2} - m_{M,1})^2} \quad (5)$$

where $m_{k,1}$ is the most prominent primitive in k -th mode in state S_1 , and $m_{k,2}$ is the most prominent primitive in k -th mode in state S_2 .

Thanks to selecting in each dimension a zero value as representation of any unclassified event, such events don't influence the distance between states.

3 Gathering the Behavioral Information

The proposed description of human activity was tested in a multimodal surveillance system for elderly. The prototype was build for studying complementary operation of wearable sensor network and building-embedded sensor network in one-to-one cooperation model (i.e. one supervised subject in his or her home). The wearable sensors accompany the subject in every activity, so the data issued are automatically labeled with the appropriate identifier. Embedded sensors data are specific for the environment and records changes made to it in result of subject activity. The system makes benefit from overlapping information from different sensing modes, and continues the monitoring in case selected data are not available (e.g. the subject is outdoor). The behavioral data is used to determine the subject's health status and commands to the smart environment. More details about the system may be found in [3].

3.1 *Decomposition of Human Shape for a Posture Detector*

An important part of the system was a motion picture-based pose detector [10], being an implementation of novel idea of an expandable fall detector. Human shape is first extracted from the environment. After calibration of the system, including background scene acquisition and expected subject appearance, subsequent video frames are segmented in order to extract the shape of the subject. Various techniques of background removal were tested including thresholding of noise in RGB and CIELab color spaces and Gaussian mixture models in HSV color space.

Contextual analysis of subsequent frames allow for extracting motion parameters for the whole body as well as for selected segments (e.g. hands or foots). Here again, many variants of thresholding of differential noise in RGB and CIELab color spaces were tested, but most reliable results were achieved with optical flow-based method. This algorithm estimates a geometrical transformation matrix best relating two considered frames. The matrix describes directions and distances of motion for each pixel in the frame.

Based on shapes and motion estimates the software builds the mobile contour of the subject, allowing for determining the pose and its temporal attributes. The size of contours has to be normalized before the pose is compared with set of patterns. The contour matching was performed in three spaces:

1. projection histograms using Manhattan metric,
2. projection histograms using Bhattachary metric, and
3. contour features provided by Snyder-Krish-Shetty method [15].

The visual system was expected to provide substantial information to the detection of subject's behavior. However, during the tests Mikrut and colleagues showed that this method hardly distinguishes 'sitting' and 'bending' positions (correct recognition rates: 58.0% and 9.6% respectively), and sometimes even confuses 'standing' with 'lying' (when a lying subject shows a remarkable vertical element) [10].

3.2 Complementary Information for Human Activity Monitoring

As demonstrated above, visual methods are not sufficiently robust to determine human activities, even belonging to assumed behavioral primitives. Therefore the visual data has to be completed by readouts from wearable sensors. In this role, we used a waistline-mounted gyroscope and wrist-mounted accelerometers. These sensors, although require independent power supply, additional subject preparation, wireless transmission, synchronization and processing of data, provide complementary information about trunk position (vertical or horizontal) and hands activity allowing for support of pose recognition and extension of primitives range by 'working' and 'walking'. The block diagram of the multimodal acquisition system for behavior investigation is presented in Figure 2.

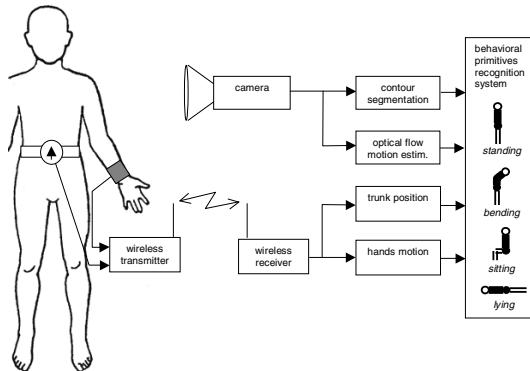


Fig. 1 Block diagram of the multimodal acquisition system for behavior investigation

4 Conditions and Results of Tests

Implementation of a prototype of assisted living surveillance system allowed for an experimental verification of the idea of using behavioral primitives for description of human activity. We performed a two step experiment having different aims:

- First, we attempted to verify how accurately the patterns specified as behavioral primitives are reproduced by a real system.

- Next, we pursuit a real live scenarios to check how the everyday life activity are decomposed into behavioral primitives.

4.1 Testing Conditions

For the first experiment we asked 17 volunteers to follow as exactly as possible the given behavioral patterns of 'standing', 'bending', 'sitting', 'lying', 'walking', and working'. Each pose was reproduced 7 times by each person. The recording was made in the laboratory in controlled lighting and uniform visual background. For the second experiment a two piece apartment was arranged in the laboratory consisting of:

- A 'bedroom' with a sofa bed, a camera (with infra red lighting) and microphones (this arrangement was also used for sleep quality studies [12], [14]), and
- A 'kitchen' with two sensorized appliances (tap and electric kettle), and a camera.

This arrangement was attributed with specific set of rules: for example no lying was allowed out of the bed, the kettle could be removed from the stand for no longer than 1 minute, the tap could be on only when the subject is close to it. etc. Four young volunteers were testing the prototype system performing actions as: lying, getting up, sitting, walking, working accordingly to 17 schedules differing by state order and duration. Each session lasts for ca. 20 minutes and aim to simulate an excerpt of everyday life. Some schedules contained events violating the rules set for the arrangement, a subset of schedules included also outdoor activities as walk, short run and rest. The behavioral patterns were collected by the sensors continuously, and the system was expected to recognize the subject status and with respect to its temporal changes to classify the behavioral pattern into one of four possible clusters: normal, suspicious, dangerous, critical.

4.2 Test Results

The accuracy of reproduction of assumed behaviora primitives in a prototype surveillance system is given in tab. 1. Comparing these data with the results of purely visual discrimination cited above allow to assess the contribution of the complementary information from the wearable sensors. The decomposition of simulated activities from everyday life into a base of behavioral primitives is presented in tab. 2. For the reason of commodity all values are normalized and given in percents. Last column of table 2 contains the value of contribution of unexplained events. This value represents how well the decomposition basis fits to actual activities.

Table 1 Results of accuracy of behavioral primitives reproduction

Primitive	Average	Standard Deviation
'standing'	97.1%	±3.3%
'bending'	67.7%	±15.8%
'sitting'	83.8%	±12.3%
'lying'	93.2%	±5.1%
'walking'	87.4%	±7.0%
'working'	91.0%	±5.8%

Table 2 Results of decomposition of activities from real live scenarios into behavioral primitives

Activity description	standing	bending	sitting	lying	walking	working	unexplained
cooking	38.1%	3.1%	5.8%	1.1%	17.8%	23.8%	10.3%
computer typing	2.1%	5.3%	37.4%	2.8%	1.3%	29.6%	21.5%
teeth brushing	26.3%	19.7%	4.8%	2.3%	6.8%	24.4%	15.7%
floor cleaning	12.7%	17.4%	5.8%	18.3%	7.1%	14.6%	24.1%

5 Discussion

We developed the idea and prototype application of behavioral primitive-based decomposition of human activity record. The set of selected primitives does not constitute an orthogonal decomposition base (in the meaning of independence of primitives), however the primitives are relatively easy for discrimination from coincidence of visual and accelerometric data. A 'standing' pose was the easiest for discrimination, but for several reasons, even for the primitives, the recognition led to ambiguous results:

- There is no parameter-based distinction between 'standing' and 'bending' pose,
- If the subject in 'bending' or 'sitting' pose is oriented face to the camera, the system confuses it with 'standing',
- If the subject in 'lying' pose shows a remarkable vertical element, he or she could be erroneously classified as 'standing',
- During the test, the poses played by volunteers showed subject-related variability.

Adding the complementary information from the wearable measurement system remarkably improved the correct recognition rate. Although 'walking' and 'working' primitives are still hardly recognized based on rhythmicity of the hands movements, that's where the optical flow analysis may help to distinguish both items. Table 2 shows that even using a simplified decomposition we could reliably guess the activity performed by the subject. Despite the part of some actions unexplained on proposed base has the highest contribution, the most important contributions of

primitives clearly reveal the principal character of the activity. This is not considered as drawback since:

- In the aspect of security of an elderly at home, the detailed information it is not required,
- The surveillance system may easily be extended by complementary sensors reporting the interaction of the subject with his or her environment (e.g. electric kettle, tap, iron etc.).

More sophisticated methods, analyzing temporal relations of motion patterns (e.g. forwards-backwards movement of a hand or alternate movements of two legs) may also bring another way to discriminate more specific actions.

Acknowledgements. The Author expresses his gratitude to Magdalena Smoleń, Zbigniew Mikrut and Przemysław Pleciak for their valuable remarks considered the research topic. This Scientific work is supported by the AGH University of Science and Technology in year 2013 as a research project No. 11.11.120.612.

References

1. Atoui, H., Telisson, D., Fayn, J., Rubel, P.: Ambient Intelligence and Pervasive Architecture Designed within the EPI-MEDICS Personal ECG Monitor. International Journal of Healthcare Information Systems and Informatics 3(4), 68–80 (2008)
2. Augustyniak, P.: Validation of automatic ECG processing management in adaptive distributed surveillance system. In: Kurzyński, M., et al. (eds.) Computer Recognition Systems 2 Springer. ASC, vol. 45, pp. 574–580. Springer, Heidelberg (2007)
3. Augustyniak, P.: Layered design of an assisted living system for disabled. In: Piętka, E., Kawa, J. (eds.) ITIB 2012. LNCS, vol. 7339, pp. 498–509. Springer, Heidelberg (2012)
4. Bousseljot, R., et al.: Telemetric ECG Diagnosis Follow-Up. Computers in Cardiology 30, 121–124 (2003)
5. Chen, X., Ho, C.T., Lim, E.T., Kyaw, T.Z.: Cellular Phone Based Online ECG Processing for Ambulatory and Continuous Detection. Computers in Cardiology 34, 653–656 (2007)
6. Chiarugi, F., et al.: Continuous ECG Monitoring in the Management of Pre-Hospital Health Emergencies. Computers in Cardiology 30, 205–208 (2003)
7. Denman, S.P., Chandran, V., Sridharan, S.: An adaptive optical flow technique for person tracking systems. Pattern Recognition Letters 28(10), 1232–1239 (2007)
8. Fayn, J., et al.: Towards New Integrated Information and Communication Infrastructures in E-Health. Examples from Cardiology, Computers in Cardiology 30, 113–116 (2003)
9. Liao, W.H., Yang, C.M.: Video-based Activity and Movement Pattern Analysis in Overnight Sleep Studies. In: Proc. 21st Pattern Recognition, ICPR (2008)
10. Mikrut, Z., Pleciak, P., Smoleń, M.: Combining pattern matching and optical flow methods in home care vision system. In: Piętka, E., Kawa, J. (eds.) ITIB 2012. LNCS, vol. 7339, pp. 537–548. Springer, Heidelberg (2012)
11. Patel, H., Wankhade, M.P.: Human Tracking in Video Surveillance. International Journal of Emerging Technology and Advanced Engineering 1(2), 1–4 (2011)
12. Puzzioli, S.: Remote Transmission and Analysis of Signals from Wearable Devices in Sleep Disorders Evaluation. Computers in Cardiology 32, 53–56 (2005)

13. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Fall Detection from Human Shape and Motion History Using Video Surveillance. In: 21st International Conference on Advanced Information Networking and Applications Workshops, AINAW 2007, vol. 2, pp. 875–880 (2007)
14. Smoleń, M., Czopek, K., Augustyniak, P.: Sleep evaluation device for home-care. In: Piętka, E., Kawa, J. (eds.) *Information Technologies in Biomedicine*. AISC, vol. 69, pp. 367–378. Springer, Heidelberg (2010)
15. Snyder, W., Krish, K., Shetty, S.: A Shape Recognition Algorithm Robust to Occlusion: Analysis and Performance Comparison (2007), <http://www.ece.ncsu.edu/imaging/Publications/2007/ShapePaper.pdf>
16. Sun, H., De Florio, V., Gui, N., Blondia, C.: Promises and Challenges of Ambient Assisted Living Systems. In: Sixth International Conference on Information Technology: New Generations, ITNG 2009, pp. 1201–1207 (2009)
17. Telisson, D., Fayn, J., Rubel, P.: Design of a Tele-Expertise Architecture Adapted to Pervasive Multi-Actor Environments. Application to eCardiology. *Computers in Cardiology* 31, 749–752 (2004)

How to Become Famous? Motives in Scientific Social Networks

Adam Matusiak and Mikołaj Morzy

Abstract. Scientometry is the discipline for measuring the success and influence of scientific work. This is usually done by analyzing scientific networks, most notably, citation networks and co-authorship networks. In our work we are taking another approach: we observe the evolution of individual scientific careers through the lens of social scientific recognition measured by the membership in program committees of conferences and editorial boards of scientific journals. Then we compare the data on program committee membership and editorial board membership with the history of scientific publications to find frequent sequences of events that lead to one's invitation to a prestigious conference or journal. We call these sequences motives and we define a few distinct classes of such motives. The large body of data harvested from the Web allows us to experimentally verify the validity and benefit of the proposed approach.

1 Introduction

Scientometrics is the area of research concerned with measuring scientific work. Many people mistakenly consider scientometrics to be the analysis of citation relationships of scientific publications or ranking of scientific journals and conferences. Also, one may find the term “scientometrics” being used to denote the measuring of individual scientists’ rankings, influence or prominence. However, these measurements are properly addressed by bibliometrics. Scientometrics, on the other hand, performs both qualitative and quantitative measurements of science. Within scientometrics one finds methods accommodated from such disciplines as data mining,

Adam Matusiak · Mikołaj Morzy
Institute of Computing Science
Poznan University of Technology
e-mail: Adam.Matusiak@cs.put.poznan.pl,
Mikolaj.Morzy@put.poznan.pl

information retrieval or machine learning. Recently, many attempts have been made to employ methods and algorithms of social network analysis and mining in scientometrics. Scientometrics also maintains its close relationships with philosophy and sociology of science, as well as the history of science [13].

We have decided to follow the social network analysis angle of research and to observe how scientific networks change over time. Contrary to previous attempts, we have not constrained ourselves to citation and co-authorship relations, but we have augmented the dataset with informations on *soft relations* between individual scientists. By soft relations we mean co-membership in program committees of scientific conferences and co-membership in editorial boards of scientific journals. We argue that this social aspect of scientific endeavor should be used as a proxy for social recognition of one's scientific value. For instance, when the program committee of a new conference is formed, it is in the best interest of conference chairs to build a program committee as famous as possible. Therefore, when deciding whom to invite to the program committee, conference chairs use their personal acquaintances to find prospective members. This process is even more expressed with scientific journals. Editorial boards of journals tend to be much more static, and the impact of the editorial board on the value and perception of the journal by the community is much greater than the influence of a program committee on the perception of a scientific conference. We draw this conclusion from our own experience of how these informal decisions are being made within the scientific community.

In our work we perform social network analysis and mining on datasets representing both publication authorship and social aspects of science environment. We are looking for meaningful and useful patterns in these merged datasets to find motives that are prominent. These sequences can be thought of as building blocks of scientific careers and using them in scientometrics provides additional insights into the informal and invisible structure of scientific community. In particular, we can observe how publishing in particular venues or co-authoring papers with particular authors influences future scientific career in terms of invitations to conference program committees or editorial boards of journals. Our goal is to discover the most reliable patterns that can both explain the social aspect of scientific development and serve as recommendations for publication venues for young scientists. Our most important contribution reported in this paper is the preparation of the dataset used to measure scientific development, identification of the most important motives and their verification in the gathered data.

The paper is organized as follows. In Section 2 we review the related work. We present the most important motives in Section 3. The dataset used in our experiments is described in Section 4 and the results of the experimental evaluation of our model are reported in Section 5. The paper concludes in Section 6 with a brief summary.

2 Related Work

The analysis of the collaboration scientific networks has been long considered to be among key applications of large scale social network analysis, as witnessed by

seminal works by Newman [8] and Barabási [1]. An interesting line of research has been proposed by Tomassini and Luthie [12] where the authors are scrutinizing the development of the collaboration network over time in search of characteristic features of the evolving network. Their work has been thoroughly grounded in a large dataset available on co-authorship and citation data. Huang et al. [5] have been pursuing a similar research trajectory modeling the evolution of the collaboration network and measuring the impact of various parameters on the shape and characteristics of the resulting network. In [10] authors attempt to map the field of Computer Science from the network perspective by clustering together similar publication venues. An interesting historical approach to scientific collaboration evolution is presented in [7], where the authors follow the development of a collaboration network from the early days until the advent of the Web. A more individualized approach is presented in [6] where the author tries to make an inference on pragmatical aspects of micro-structures that appear among collaborating individuals. However, previous work differs significantly from what we are doing, since we are employing the data on program committee and editorial board membership. As we have previously stated, these data represent sociological aspects of the scientific endeavor and they allow us to measure the publicly perceived reputation of individual scientists, that may stem from reasons other than pure scientific merits. To the best of our knowledge this is the first attempt to observe how one's publication career is influenced by social milieu and recognition from the peers.

Another area strongly related to our current research is the scientometrics. It is a well established research discipline with its core ideas described in works such as [13], [4] and [11]. Scientometrics uses not only the analysis of co-authorship and citation, but reaches to the philosophy of science [9] and sociology of science [2]. Our work extends previous attempts by incorporating data which represents social aspects of scientific environment, yet does not come from surveys (which can and often are biased), but from hard data.

3 Motives

Let us now introduce basic notation used throughout the remainder of the paper. We are concerned with the discovery of sequences of atomic events, such as publishing a scientific paper, being invited into a program committee, or being added to the editorial board of a journal. We are trying to find the most pronounced sequences of events, and we will refer to frequent sequences as *motives*. Let us use the following notation:

- $E_t(a_i, e_j)$ is a predicate stating that the author a_i belonged to the program committee of the venue e_j at the time t^1 ,

¹ We deliberately use e_j to denote a publication venue (conference or journal) because we treat both the fact of publishing a paper and the fact of belonging to a program committee or editorial board as certain types of *events*.

- $P_t(a_i a_j a_k \dots, e_j)$ is a predicate stating that the authors a_i, a_j, a_k, \dots published in the venue e_j at the time t ,
- $C_t(a_i, a_j)$ is a predicate stating that authors a_i and a_j co-authored a paper at the time t .

Using these atomic events as building blocks we can introduce the basic motives for scientific career development. We will further refer to the left hand side of the motive as the *antecedent* of the motive, and the right hand side of the motive as the *consequent* of the motive.

1. *Introduction by publication:* $P_t(a_i, e_j) \rightarrow E_{t'}(a_i, e_j), t < t'$ The author a_i first publishes in a venue, and as the result the author is invited to join the program committee or editorial board of the venue. An important constraint is that the publication of a paper must precede the invitation to the committee or board. This motive represents the most frequent and common way of replenishing the ranks of program committees.
2. *Introduction by cloning:* $E_t(a_i, e_j) \rightarrow E_{t'}(a_i, e_j), t < t'$ This motive reflects the natural evolution of program committees and editorial boards, where the list of members is simply cloned from the preceding year. One may think of this motive as the typical static transition of a program committee or editorial board from one year to another, with scientists simply transferred between consecutive publication venues.
3. *Publication by repetition:* $C_t(a_h, a_j) \rightarrow C_{t'}(a_h, a_j), t < t'$ This simple motive measures the willingness of people to cooperate in the future after producing a joint research. In itself this motive does not contribute to the development of scientific career, but in connection with the next motive it may be important to realize the instances of publication by repetition.
4. *Introduction by co-publication:* $C_t(a_h, a_i) \wedge P_t(a_h a_i, e_j) \rightarrow E_{t'}(a_h, e_j) \wedge E_{t'}(a_i, e_j), t < t'$ This motive measures the probability that a joint publication between authors a_h and a_i in the venue e_j increases their chances of being invited to serve in the committee for this venue. In other words, instances of this motive represent successful development of a scientific career as a result of co-operation with a (possibly famous and recognized) scientist.
5. *Publication after introduction:* $E_t(a_h, e_j) \wedge E_t(a_i, e_j) \rightarrow P_{t'}(a_h a_i, e_j), t < t'$ The final motive reflects the tendency of scientists to publish their work (especially joint work) in publication venues for which they serve in program committees and editorial boards. This is understandable since one of the main reasons of including a particular scientist in a program committee or editorial board is the possibility of gaining more submissions using social environment of the invited individual.

Of course, this list is far from complete, it is an initial attempt to recognize most typical motives that constitute scientific careers of individuals. For each of the presented motives we compute the number of occurrences of the antecedent, the number of occurrences of the consequent, given that the antecedent has also occurred, and the conditional probability of the motive. These measures have been computed both for the entire dataset (denoted by DBLP), and for the dataset constrained to

Polish scientists only (denoted by OPI) in order to compare how the motives differ between a localized subset and the entire dataset. Table 1 contains the results.

Table 1 Frequency and confidence of motives

motive	DBLP			OPI			
	total	match	prob.	total	match	prob.	ratio
M_1	996 890	29 863	3%	16 268	833	5.12%	1.71
M_2	214 542	33 850	15.78%	6738	921	13.67%	0.87
M_3	3 578 579	649 114	18.14%	67 454	13 226	19.61%	1.08
M_4	3 578 579	71 027	1.98%	67 454	1753	2.60%	1.31
M_5	12 532 066	13 140	0.10%	303 217	332	0.11%	1.04

There is very little difference in the presence of motives between the DBLP and the OPI datasets. Furthermore, we see a strong incentive to continue previous collaboration (almost 20% of pairs of co-authors produced at least one more joint publication in the future). Committees are partially built from previous members as the motive “Introduction by simple cloning” is quite common. On the other hand, we do not see a significant influx of new members into the committees and editorial boards from amongst the authors, the motive “Introduction by publication” does not seem to be popular.

4 Dataset

We have harvested a large dataset consisting of scientific papers’ co-authorship, augmented with the information on the submissions of papers to conferences and journals, as well as information on the participation of individual scientists in program committees of conferences and editorial boards of journals. Following the terminology introduced by Granovetter [3] we refer to the relationships of co-authorship as *strong ties*, and we treat the relationships of co-membership in a program committee or editorial board as *weak ties*. The rationale between this distinction is such that the co-authorship of papers implicates friendship, joint effort and first-hand acquaintance, while sitting on the same committee might, but as well might not, signify direct acquaintance between scientists. The primary source of data used in our research was provided by DBPL², a large online library of computer science publications. DBLP provided data on authors’ names, publications, publication venues and co-authorship of publications. We have also crawled the WikiCFP³ website for all “Call for Paper” announcements. We have parsed these documents to identify scientists listed in program committees of conferences. For journal editorial boards

² <http://dblp.uni-trier.de>

³ <http://www.wikicfp.com/cfp>

we have crawled journal webpages for lists of editorial board members. The final result was the dataset with the following characteristics:

Table 2 Dataset characteristics

Entity	Quantity	Entity	Quantity
scientist	1 025 692	publication	1 759 452
publication-author link	4 722 415	book/journal	32 450
committee-member link	410 044	committee	14 531
event	1 773 983	conference	6146

We have also created a constrained dataset with very detailed and high-quality data from the OPI database⁴. We have discovered 156 595 unique profiles of scientists, 11.24% of whom were matched with the DBLP dataset (which means that Polish computer science is over-represented in OPI database). Our dataset focuses on computer science and other closely related fields. Thus, although our experimental evaluation of motives is valid for computer science, other domains of science may be characterized by significantly different publication culture.

5 Experiments

In order to perform a more detailed scrutiny of each motive, we have computed the time-series of each motive and the analysis of the time span between the antecedent and consequent of each pattern. Below we present the results along with a short commentary. Please bear in mind that the dataset has been collected at the very beginning of 2012, so the last data points represent the year 2011. This may influence the frequencies and confidences of motives at the end of each scale, close to year 2011, because for many motives the antecedent is present in the database, but the consequent has not happened yet. In all the figures “Total” corresponds to the entire dataset, while “OPI” represents the dataset constrained only to the scientists present in the OPI database. Timespan is measured in years and all reports on the timespan have been normalized to the range of $\langle 0, 1 \rangle$.

As can be easily noticed in Fig. 1, this motive is decreasing in frequency. The left hand side figure presents the frequency of motive appearance throughout the years, while the right hand side figure presents the normalized probability of the time span of the motive (i.e. how probable it is that the consequent of the motive follows the antecedent of the motive in a given number of years). We see that it becomes harder and harder to construct a career in the scientific community only by publishing in a conference or a journal, because only 7% of authors will be invited

⁴ OPI is the scientific unit of the Ministry of Science and Higher Education responsible for managing information on the entire Polish science

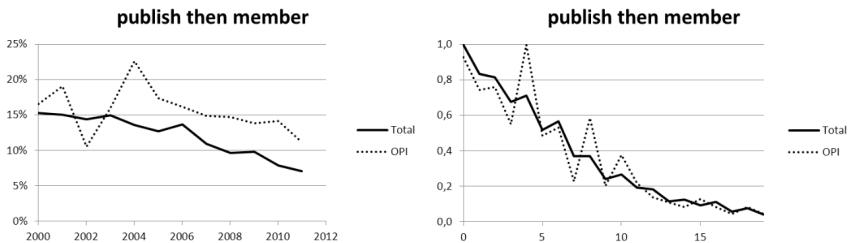


Fig. 1 The frequency and the time span of the pattern *Introduction by publication*

to join the committee or an editorial board of their respective publication venue. Secondly, an invitation usually happens directly after publication, the probability of invitation into the committee as the result of a publication decreases linearly with the time passed from the publication date.

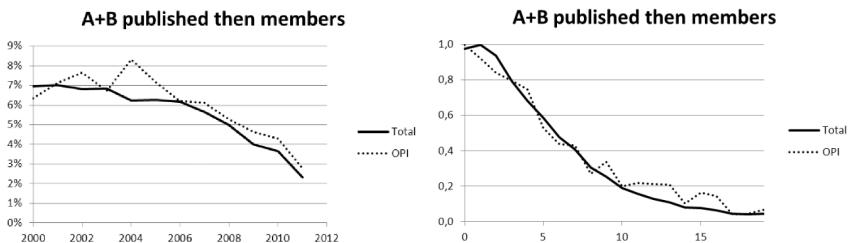


Fig. 2 The frequency and the time span of the pattern *Introduction by co-publication*

Introduction by co-publication was a popular motive until 2006 (Fig. 2), but since that year we see a steady decrease in the frequency of this motive. We see this behavior both in the full dataset and in the dataset constrained to Polish computer science. The reasons for the decline in co-publication remains unknown and requires further investigation. The time span distribution is similar to the first motive, an invitation is most likely within one year of the publication.

In Fig. 3 we are observing a stable distribution of the motive “Introduction by cloning”. There is a 50% chance that a member of the program committee or editorial board will continue to serve in the subsequent year. An interesting phenomenon is visible in the time span distribution for this motive. The cloning of a program committee or editorial board happens almost exclusively within one year (a large spike around value 1 in the figure). This suggests that once a person quits a program committee or editorial board, the probability of return is extremely small. This cloning technique is ubiquitous and, in our opinion, it hurts the dynamics of the science environment because it effectively blocks access of junior researchers to program committees and editorial boards.

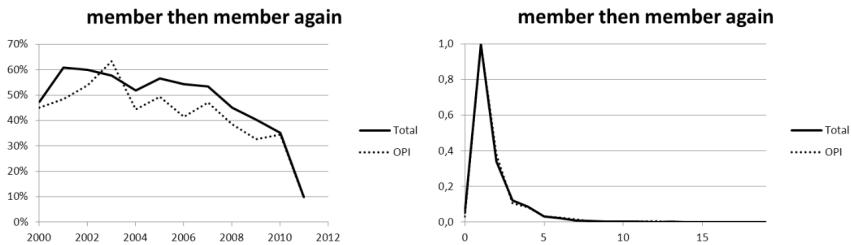


Fig. 3 The frequency and the time span of the pattern *Introduction by cloning*

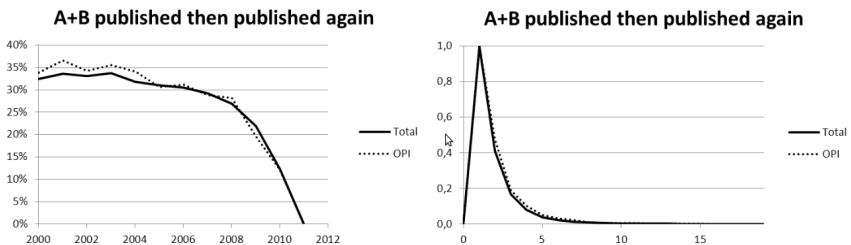


Fig. 4 The frequency and the time span of the pattern *Publication by repetition*

In order to understand the frequency distribution of the motive “Publication by repetition” one must remember that our dataset might not contain enough data points representing events from 2011. This is our only explanation for the sudden drop in frequency displayed in Fig. 4. Otherwise the distribution is stable and we see that on average one in three pairs of collaborators will publish joint research again in the future. Interestingly, the time span distribution suggests that there is a very strong inclination towards continuing joint research in short time spans. Most probably, what we are seeing here are multiple joint publications reporting on a single research project that contribute so strongly to this particular shape of the distribution.

The fifth motive “Publication after introduction” (Fig. 5) might seem counter-intuitive. This motive represents a publication in a venue after being invited to the program committee or an editorial board. One may think of this motive as representing the social role of program committees which allow researchers to strengthen their professional ties and start research projects. Indeed, the frequency of this motive is very low as compared to other motives. Furthermore, it might be the case that the two researchers share their affiliation which predominantly drives their co-operation and that there is no causal relationship between the antecedent and the consequent of the motive. We note that this motive requires further investigation.

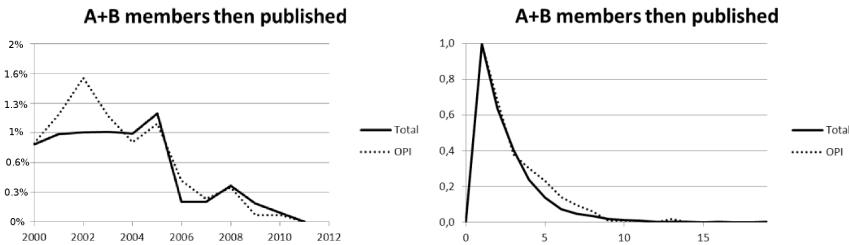


Fig. 5 The frequency and the time span of the pattern *Publication after introduction*

In order to investigate whether this closed form of PCs is really hurting young scientists we have measured the dynamics of the change of the cardinality of PCs. The results are presented in Fig. 6. As we can see (left figure), the number of conferences and workshops grows quickly. However, the structure of a PC remains stable with the average number of members being around 18. However, there is a great variability between conferences and workshops, with few very large conferences and many more small conferences and workshops. This variability is reflected by high values of the standard deviation. Because the cardinality of PCs is not normally distributed, we use additional measures of median and median absolute deviation which are more appropriate for such distributions. We see that both measures remain very stable throughout the years (right figure). Thus, we conclude that despite the enormous growth of the number of conferences and workshops, their cardinality remains remarkably stable and the ubiquitous copying of one year's PC to the next year is indeed hurting younger scientists.

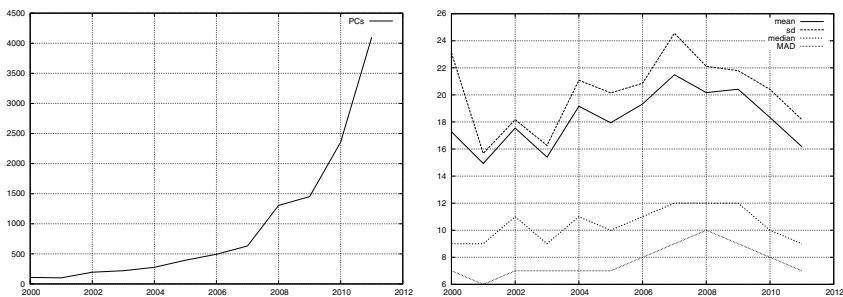


Fig. 6 Dynamics of PC's cardinality change

6 Conclusions

In this paper we have presented the new model of motives for the analysis of scientific cooperation networks. We want to extend this simple model and combine it with Bayesian network techniques to build a predictive model capable of suggesting possible development of young scientists' careers resulting from particular publication choices. Such network could be used as a recommendation engine that would suggest conferences with the greatest possible payoff in terms of scientific social recognition. It is obvious that for each discipline there exists a set of top tier conferences for which no recommendations are necessary. Nevertheless, young researchers are faced with a huge number of secondary publication venues and it is not clear which ones to choose to maximize the development of scientific career, given limited research funds and publication possibilities.

References

1. Barabási, A., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications* 311(3-4), 590–614 (2002)
2. Ben-David, J., Sullivan, T.: Sociology of science. *Annual Review of Sociology* 1, 203–222 (1975)
3. Granovetter, M.: The strength of weak ties. *Am. J of Sociology*, 1360–1380 (1973)
4. Hou, H., Kretschmer, H., Liu, Z.: The structure of scientific collaboration networks in scientometrics. *Scientometrics* 75(2), 189–202 (2008)
5. Huang, J., Zhuang, Z., Li, J., Giles, C.L.: Collaboration over time: characterizing and modeling network evolution. In: Proc. of the Int. Conference on Web Search and Web Data Mining, WSDM 2008, pp. 107–116. ACM, New York (2008)
6. Melin, G.: Pragmatism and self-organization: research collaboration on the individual level. *Research policy* 29(1), 31–40 (2000)
7. Moody, J.: The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review* 69(2), 213–238 (2004)
8. Newman, M.E.J.: Scientific collaboration networks. network construction and fundamental results. *Phys. Rev. E* 64, 016,131 (2001)
9. Papineau, D.: Philosophy of science. Wiley Online Library (2007)
10. Pham, M.C., Klamma, R.: The structure of the computer science knowledge network. In: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 17–24 (2010)
11. Schubert, A.: The web of scientometrics. *Scientometrics* 53(1), 3–20 (2002)
12. Tomassini, M., Luthi, L.: Empirical analysis of the evolution of a scientific collaboration network. *Physica A: Statistical Mechanics and its Applications* 385(2), 750–764 (2007)
13. Van Raan, A.: Scientometrics: State-of-the-art. *Scientometrics* 38(1), 205–218 (1997)

AdaBoost for Parking Lot Occupation Detection

Radovan Fusek, Karel Mozdřeň, Milan Šurkala, and Eduard Sojka

Abstract. Parking lot occupation detection using vision systems is a very important task. Many systems use different sensors and their combinations to find out whether the parking lot or space is occupied or not. Using CCTV systems makes it possible to monitor great areas without a need of many sensors. In this paper, we present a method that uses the boosting algorithm for car detection on particular parking spaces and shifting the image to obtain a probability function of car appearance. Using the model of parking lot, we achieve occupancy of each parking space. We also experimented with the detector that is based on the histogram of oriented gradients (HOG) with a support vector machine (SVM). Nevertheless, we found some drawbacks of this detector that we describe in experiments. On the grounds of these drawbacks, we decided to use the AdaBoost based detector.

1 Introduction

Traffic information systems rely on multiple information sources. It is necessary to acquire as much data as possible, but we also require high reliability of this data. Parking lot occupation information systems are useful for intelligent car assistants for on-line planning and finding of spaces for parking.

In some parking lot occupation detection systems, many different sensors are used and it considerably increases the price of whole system. Recently, these sensors are replaced with CCTV systems in many fields. The motivation for this is that one camera placed on parking lot can monitor all parking spaces at once.

Radovan Fusek · Karel Mozdřeň · Milan Šurkala · Eduard Sojka
Technical University of Ostrava, FEECS, Department of Computer Science,
17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic
e-mail: {radovan.fusek, karel.mozdren, milan.surkala,
eduard.sojka}@vsb.cz

The problem of parking lot detection system is to find out occupation of parking spaces. The systems that are available can be divided into two categories, car based systems and parking space based systems. The car based systems are based on finding the cars in the images. The parking space based systems depend on their state with respect to their particular area, which is often combined with the background subtraction models.

In our approach, we work with the images from a camera and we use the AdaBoost based algorithm to find cars in the specific parking space. If it is found, then the parking space is marked as occupied. Otherwise, it is marked as free. Boosting algorithms are quite fast, but there might be many parts of the image that can be similar to cars. We use only a part of car image for detection and this way we try to minimize the error of false detection. Therefore, our method is a combination of car and space driven methods.

We also shift the detection area in the image to compute probability function. The function is computed as a sum of successful car detections. According to this and a preset threshold, we mark the parking space as occupied or not. This makes our approach less false detection prone.

In this paper, we describe our approach in detail. We provide the readers with information on the boosting algorithms and the problem of classification. Then we show the results of our approach and discuss them in the conclusion.

2 Releted Work

Many methods for object detection and recognition using machine learning methods were developed recently [3, 8, 10]. The object detection systems can be divided into several classes: feature based, model based, background subtraction based, etc. Related work based on car detection can be also divided into the two main classes [4]: explicit and implicit.

The explicit models [5, 7, 9] use generic models of cars. It is represented as an 2D or 3D model of the car. The models are also often decomposed to sub models and are detected separately according to their features. The combination of these features is then used for the detection of whole car. The most used features are based on rectangular car borders, windshields, intensity of color, or intensity of shades. The car is detected if enough evidence of car presence is found.

Implicit modeling is based on appearance [1, 3, 6]. It uses sample images to describe the model of the car. Classifiers or statistical tools are used in this modeling for classification of the input samples according to the extracted features. For example, the detection methods that are based on the support vector machine (SVM) were presented in [2, 8]. The algorithms for detection are based on finding (matching) the car models in the whole input image. Some methods are focused on contextual knowledge [7, 13]. In this case, the cars are detected only on roads or parking lots.

The detectors that are based on the histogram of oriented gradients (HOG) with SVM are also capable of this detection task [2]. Nevertheless, this detector had a

high number of false positive detections in noisy images. That is the reason why we use the algorithm that are based on the Haar-like features with AdaBoost. We show in our experiments that the proposed detector achieves better results in noise images than the detector that is based on HOG with SVM.

3 Boosting, AdaBoost

Our training data obtained from the parking space can be used with many classifiers. We use the algorithm proposed by Viola and Jones [11, 12] due to its success in many detection tasks and due to the reasons that was described above. Viola and Jones used Haar-like features, integral images, AdaBoost and cascade structure in their framework.

Haar-like features consist of white and black regions and the features are defined as difference of the sum of pixels between the regions. The response to the Haar-like features is computed very quickly using the image representation called integral images. The features can be used as weak classifiers. Generally, many variations of these features can be formulated.

AdaBoost is used to select the most useful and relevant classifiers (features). The main idea of AdaBoost algorithm is to create a strong classifier as a linear combination of weighted simple weak classifiers. Single Haar-like feature can represent a simple weak classifier. The input for AdaBoost is a training set $S = (x_1, y_1), \dots, (x_m, y_m)$, where x_i represent sample (in our case cars) and y_i represent possible evaluation (positive samples are marked as 1, negative samples has 0). The algorithm works in an iterative way and it is changing the importance of data in each its step. Finally, AdaBoost choose effecient features ("weak" classifiers) according to the lowest error and it creates "strong" classifier using a linear combination of weak classifiers.

Viola and Jones object detection framework uses cascade structures to achieve a better classifier speed. The cascade consists of several stages. The stages consist of several simpler classifiers. The main idea of cascade is the rejection of negative candidates in the early stages of cascade. The candidate that successfully passes all phases of cascade is marked as a positive detection (in our case as the car).

4 Our Approach – Proposed Method

Our approach consists of several steps. These steps are shown in the diagram in Figure 1. In the first step, we obtain the image from the camera. The location of our camera is on the roof of the near building that is above the parking lot. Sample images obtained from this camera are shown in Figure 2. In the second step, we get the coordinates of the parking spaces. Each parking space is defined by four coordinates and we manually draw a map of parking spaces. The map of parking spaces is

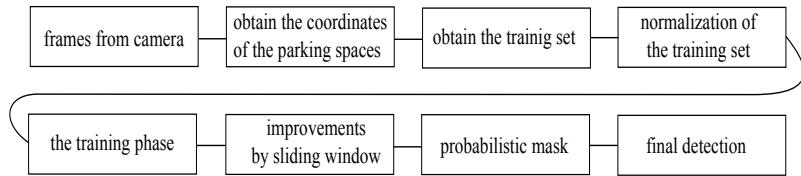


Fig. 1 The diagram containing the basic steps of our approach

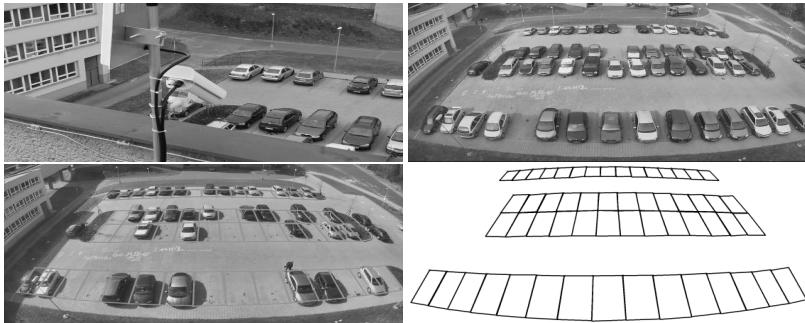


Fig. 2 The images from our parking lot

shown in Figure 2. Finally, the coordinates are obtained for each parking space from this map. After obtaining the coordinates, we can process images from each parking space separately. These images are normalized for the training phase. The normalization includes an important preprocessing steps: a perspective transformation, cut off the background and size normalization. We use a normalized images as an input to the training phase. After the training phase, we obtain our classifier. We use this classifier in the detection phase. In the next section, we suggest the following steps to improve detection in our application of parking space.

4.1 *Improvements of Detection Rate*

In order to achieve higher detection rate, we propose several techniques that should increase rate of correct car detection and concurrently decrease false positive rate. The first technique is called *window shift*. We search for car not only in the window located in a position of parking space, but we also slightly shift this window in 8 additional directions (right, left, top, bottom, right-top, left-top, right-bottom and left-bottom). The detection algorithm is carried out in all these positions. The number of car detections in one area (9 windows) tells us about the probability of occupancy. If the number is high and it is close to 9, we can mark this space as

an occupied one. On the other hand, if the number is close to zero, there is a high probability that the detection was only accidental. In order to mark these position as correctly as possible, we use a probabilistic function which will be presented later. It is obvious that this approach reduces the situations of accidental detections of cars in empty spaces and detections of empty spaces in occupied spaces.

The second technique is called the *reduction of window*. Because of shifting the window, higher probability of false detection in empty spaces can emerge (despite that it mostly reduces this problem). It is caused by interference of neighboring car to detected area. Therefore, we slightly lower the size of searching window by 5 – 10%. Because of that, the shifted windows do not cover so much the neighboring parking spaces and it reduces the false positive rate. Concurrently, it still preserves the advantage of shifting (probability of detection of wrong parked cars and higher probability of correct result because of the higher number of detections).

The third technique is optional and can be eliminated by better marking of parking spaces. The problem lies in strong perspective distortion of parking spaces near the edge of the image. Therefore, we should not mark the parking spaces in the ground level because a large portion of the car could be depicted in a neighboring parking space or even out of the parking lot. It is more appropriate to mark the parking space approximately in the height of headlamps. On the other hand, it is more simple to mark the spaces in the ground level and adjust the searching window by our third technique of automatic *perspective correction*. This correction should increase the car detection rate in the edges of frame but is also could suffer from the problem of detection of boundary stones and shadows as cars. Therefore, the detector should be learned for such negative images, otherwise, many of empty spaces in the edges could be detected as cars. Our classifier was not trained for such cases yet, so we do not use this improvement in our tests. Preliminary tests showed better results in moderately distorted spaces but higher false positive rates were achieved in extreme edges and this should be addressed in future work.

As mentioned before, the output from the shifting improvement gives us a probabilistic model that increases the reliability of parking space occupancy detection. Each time car is detected in shifted area, a Gaussian function is cumulated in probability image. Once it is done for all parking spaces, we get the probability that gives us information about presence of the car in each part of the image (Fig. 3). Each parking space is then processed separately and a threshold is set to distinguish whether the parking space is occupied or not.

Many detectors use the whole image as an input. These detectors process the input image in all positions, scale etc. Such approaches are computationally demanding. On the other hand, these approaches allow detection of objects at various locations and sizes. The motivation for the use of our approach with our proposed improvements is to avoid processing the whole image in all different scales and positions in order to accelerate the detection phase of the algorithm while maintaining satisfactory detection results.

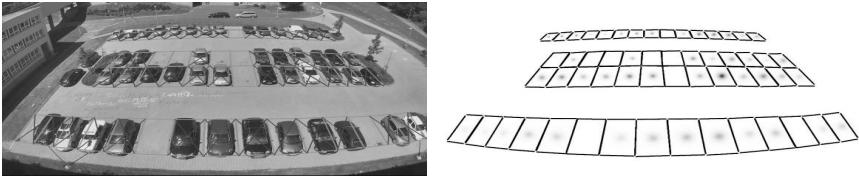


Fig. 3 The accumulated probabilities of car presence over the whole image

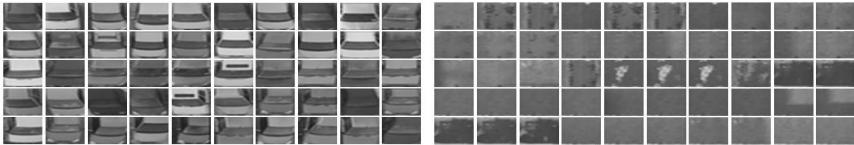


Fig. 4 The example of positive (left) and negative (right) training samples

5 Experiments

5.1 Training Set

We created a set of training images that contains 4500 positive and 4500 negative samples. The examples of these positive and negative samples are shown in Fig. 4. The positive set consists of cars from the first row (the cars closest to the camera). The cars from the first row are slightly similar to the cars from the back rows (the cars differ mainly in their size between the front and back rows). Furthermore, the samples are not affected by the camera distortion in the first row. The cars in our set were obtained with different climatic and lighting conditions for better detection properties. Moreover, these samples have good image quality and images taken from the first row have the potential to create a data set of a car in the future. Such images could be used in other detection applications. Each positive sample was resized to 20×24 pixels for the training phase.

The negative set of training images was taken from our parking spaces. Each parking space (without cars) was used to the training set. We created this negative set with different climatic and lighting conditions for better detection properties as in obtaining the sample of cars.

5.2 Real World Tests

Before continuing to experiments with the proposed detection method, we regard as desirable to describe the selection of this approach instead of detector that is based on the HOG descriptors with SVM. We experimented with the parameters of HOG based detectors and we suggested the optimal configuration (HOG_1). Each positive

Table 1 The detection accuracy of proposed algorithm and the HOG based detector in noisy images

Approach	Cars	Det. Cars	Rate	Empty sp.	Det. spaces	Rate	Accuracy
HOG_1	18	17	94.4 %	430	179	41.6 %	43.8 %
Proposed	18	14	77.7 %	430	425	98.8 %	94.2 %

Table 2 The detection accuracy of algorithm and its enhancements

Move	Resize	Cars	Det. Cars	Rate	Empty sp.	Det. spaces	Rate	Accuracy
-	-	591	566	95.8 %	417	379	90.9 %	93.8 %
yes	-	591	580	98.1 %	417	370	88.7 %	94.2 %
yes	-2 %	591	578	97.8 %	417	378	90.6 %	94.8 %
yes	-4 %	591	578	97.8 %	417	381	91.4 %	95.1 %
yes	-6 %	591	576	97.6 %	417	387	92.8 %	95.5 %
yes	-8 %	591	563	95.3 %	417	390	93.5 %	94.5 %
yes	-10 %	591	563	95.3 %	417	397	95.2 %	95.2 %
yes	-12 %	591	548	92.7 %	417	401	96.1 %	94.1 %
yes	-14 %	591	550	93.1 %	417	402	96.4 %	94.4 %

and negative sample was resized to 96×96 pixels for the training phase of the HOG detector. The configuration of HOG_1 was designed with the size of block = 16×16 , size of cell = 8×8 , horizontal step size = 8, number of bins = 4. This configuration achieved high detection rates in day illumination. Nevertheless, the HOG based detector (with SVM) had the high number of false positive detections in night illumination (in images with noise) than the presented detector. The artifacts that are created by noise have a negative effect to efficient computation of HOG descriptors and the HOG based detector often detects the spaces as occupied although the cars are not in these spaces. The detection rates of detectors (without improvements that we proposed in section 4.1) in noisy images are shown in Table 1.

For the above reasons, we decided to continue to test the presented approach in next experiments only, because we needed a robust detector that is able to detect occupied and vacant spaces in parking lots under day and night illumination with satisfactory results in both illuminations. In order to prove the usefulness of our algorithm in day illumination, we picked up 18 images of park lot in different times of a day. We chose not only images in the cloudy weather, but we also used images in a strong sunlight which causes strong shadows. 9 overcast and 9 sunlight images were chosen, each of them had 56 parking spaces. In total, 1008 parking spaces were examined, 591 were occupied and 417 of them were empty. We show the detection rates with the basic algorithm, algorithm enhanced by moving detection window and shrinking its size in Table 2. This table shows detection results for all test images together (cloudy and sunny weather).

Even the basic algorithm without the enhancements gives nice results with 95.8% correctly detected cars and 90.9% empty spaces. In total, 93.8% of parking spaces



Fig. 5 The results of detection

were marked correctly. Both numbers can be further improved. If we use the moving detector window, more than 98% of cars were correctly detected, but the detection of empty spaces decreased to less than 89%, even though the accuracy slightly raised. The improvement of car detection rate was achieved by searching for the car multiple times and we detect 14 more cars from 25 previously undetected. Therefore, the car can be detected multiple times even in situations when it was not detected by the basic algorithm at all. Moreover, we can detect the wrongly parked car more easily. The drawback is that we can also detect the wrongly parked cars in empty spaces by shifting the detector window. In order to improve this issue, we were shrinking the size of detector window. It should still detect the wrongly parked cars in the processed parking space but also should not cover the big part of car from the neighboring parking space. It is obvious that decreasing the size of detector window slowly reduces the car detection rate and slowly increases the empty space detection. Many experiments proved that best results are achieved by shrinking the window by -6 % -8 %. In such cases, the car detection rate is still higher than 97.5 % and empty spaces are correctly detected in slightly more than 95 %.

Table 3 shows the difference between cloudy weather (without shadows) and sunny weather (with problematic shadows). The moving detector window is a significant help in overcast images. The car detection rate rises up to 99 % and we achieved about 96 % to 98 % empty space detection rate with the moving window and the small shrinking factor. The accuracy is approx. 97.5 % to 98 % with the shrinking factor from -2 % to -10 %.

On the other hand, the sunny images can be problematic because of strong shadows casted by the neighboring cars to the empty spaces. The car detection rate is still high even in these difficult conditions with 96 % to 98 % using the shrinking factor from -0 % to -6 %. The smaller windows than that quickly reduces the car detection rate. Shadows are a real problem because the empty parking space detection is reduced to 85 % to 88 % with the same parameters and quickly rises up to 95 % with smaller windows. The goal is to choose such shrinking factor that has still the high car detection rate and also has not the very low empty parking space rate. In

Table 3 The detection accuracy of algorithm and its enhancements for different conditions

		Cloudy Weather			Sunny Weather		
Move	Resize	Det. Cars	Det. Spaces	Accuracy	Det. Cars	Det. Spaces	Accuracy
-	-	95.3 %	93.2 %	94.4 %	96.3 %	88.6 %	93.1 %
yes	-	98.0 %	91.7 %	95.4 %	98.3 %	85.8 %	93.1 %
yes	-2 %	99.0 %	96.6 %	98.0 %	96.6 %	84.8 %	91.7 %
yes	-4 %	98.3 %	96.1 %	97.4 %	97.3 %	86.7 %	92.9 %
yes	-6 %	99.0 %	97.6 %	98.4 %	95.9 %	88.1 %	92.7 %
yes	-8 %	97.0 %	98.1 %	97.4 %	93.5 %	89.1 %	91.7 %
yes	-10 %	97.0 %	98.5 %	97.6 %	93.5 %	91.9 %	92.9 %
yes	-12 %	94.6 %	97.6 %	95.8 %	90.8 %	94.8 %	92.5 %
yes	-14 %	95.3 %	98.1 %	96.4 %	90.8 %	94.8 %	92.5 %

**Fig. 6** The results of detection in another parking lot

future work, we are going to try to improve the empty parking space rate in sunny weather by a shadow removal. The result of detections are shown in Figure 5.

We tried to apply our algorithm to the another parking lot with the same classifier. The results are presented in Fig. 6. It is obvious that the trained classifier is sufficient even for the another parking lots. In all three testing images, we achieved 100.0 % detection rate, but we need to say that this parking lot is more simple for the detection algorithm than the former one.

6 Conclusion

We presented the AdaBoost based algorithm for the detection of occupancy of parking spaces provided with the few improvements ensuring the higher robustness of our method. Our test showed that this method provided the high detection rate of about 98 % in cloudy weather a 93 % in strong sunlight. In future work, we will focus on improvement in sunny images, probably with the shadow removal method.

Acknowledgements. This work was supported by the SGS in VSB Technical University of Ostrava, Czech Republic, under the grant No. SP2013/185.

References

1. Bileschi, S.M., Leung, B., Rifkin, R.M.: Component-based car detection. *Street Scene Images*. S.M. Thesis, EECS, MIT (2004)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893 (June 2005)
3. Heisele, B., Riskov, I., Morgenstern, C.: Components for object detection and identification. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds.) *Toward Category-Level Object Recognition*. LNCS, vol. 4170, pp. 225–237. Springer, Heidelberg (2006)
4. Hinz, S.: Detection and counting of cars in aerial images. In: Proceedings. 2003 International Conference on Image Processing, ICIP 2003, vol. 3, pp.III- 997-III-1000 (September 2003)
5. Hinz, S., Schlosser, C., Reitberger, J.: Automatic car detection in high resolution urban scenes based on an adaptive 3d-model. In: 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, pp. 167–171 (May 2003)
6. Leibe, B., Leonardis, A., Schiele, B.: Combined object categorization and segmentation with an implicit shape model. In: ECCV Workshop on Statistical Learning in Computer Vision, pp. 17–32 (2004)
7. Moon, H., Chellappa, R., Rosenfeld, A.: Performance analysis of a simple vehicle detection algorithm. *Image and Vision Computing* 20(1), 1–13 (2002)
8. Papageorgiou, C., Poggio, T.: A trainable system for object detection. *Int. J. Comput. Vision* 38(1), 15–33 (2000)
9. Ram, T.Z., Zhao, T., Nevatia, R.: Car detection in low resolution aerial images. In: *Image and Vision Computing*, pp. 710–717 (2001)
10. Schneiderman, H., Kanade, T.: A statistical method for 3d object detection applied to faces and cars. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 746–751 (2000)
11. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, pp. I-511 – I-518 (2001)
12. Viola, P., Jones, M.J., Snow, D.: Detecting pedestrians using patterns of motion and appearance. In: Proceedings of Ninth IEEE International Conference on Computer Vision, vol. 2, pp. 734–741 (October 2003)
13. Zhao, T., Nevatia, R.: Car detection in low resolution aerial images. *Image and Vision Computing* 21(8), 693–703 (2003)

Blink Detection Based on the Weighted Gradient Descriptor

Krystian Radlak and Bogdan Smolka

Abstract. In this paper some modifications of the eye blink detection method based on the weighted gradients are presented. We propose some novel techniques of denoising of the obtained waveforms and fully automated zero-crossing detection capable to detect eye blinks. These modifications were tested on two different databases. The evaluation of results show that the introduced modifications improve performance of the proposed detection framework, in which the pixels of each video frame are divided into two groups according to the direction and magnitude of the hybrid gradient vectors. The distance between their centers of gravity is used for the determination of the eye movement characteristics. The proposed technique can also be used for the analysis of eye movements and can be utilized in systems which are monitoring human fatigue, drowsiness and emotional states.

1 Introduction

The eye movements analysis is a very active research area related to fatigue, drowsiness or human cognition [1, 17]. Variability in eye blink rate is subject of research of psychologists, ergonomists and engineers, but manual analysis of blinks is very tedious and time consuming.

The study of a correlation between eye blinking dynamics and human behavior in the real world applications are of high importance and potential, as eye movements can reliably reflect the human emotional conditions [3, 13]. The evaluation and recognition of eyelid movements can limit the accidents caused by falling asleep and help report the efficiency and concentration at work. It can also be used in the human-machine interface devices and can be applied as an indicator of some typical signaling method for people with limited upper body mobility [8].

Krystian Radlak · Bogdan Smolka
Silesian University of Technology
e-mail: {krystian.radlak, bogdan.smolka}@polsl.pl

The eye blinking is also the subject of research conducted by psychologists. One of the most interesting research fields in behavioral science is the analysis of blinking rate treated as a deception indicator. Elevated blink rates are observed in subjects who try to mask their true emotions [15], and a decreased blink frequency is typical for people who are not telling the truth, followed with an increased blinking afterward [9].

Despite the great interest in the analysis of human blink patterns, effective and accurate computer vision algorithms for the eye movement analysis are still under development. In our recent work [16], we presented a brief overview of existing methods and presented a novel blink detection algorithm, which we call the Weighted Gradient Descriptor (WGT). The accuracy of the algorithm was evaluated on 5 video sequences registered in our laboratory with the high speed camera and on the ZJU Eyeblink Database [12], which was used as a benchmark in some relevant papers [5–7, 12, 18, 20].

Pan et al. [7, 12] introduced using the ZJU database, an undirected Conditional Random Field framework (CRF), incorporated with a discriminative measure of eye states. In their approach, a three-state set for eyes: open α , closed γ , ambiguous β are defined and the eye blink was modeled as a state change pattern of $\alpha \rightarrow \beta \rightarrow \gamma \rightarrow \beta \rightarrow \alpha$. CRF allows to estimate the probability of a transition between a given state and offer several advantages over hidden Markov models.

Divjak and Bischof [5, 6] used the normal flow and OpenCV's implementation of the Adaboost and Lukas-Kanade feature tracker to detect and track face and eyes position using the ZJU database. The normal flow technique is utilized to describe four states: *Open*, *Closing*, *Closed* and *Opening*, that allow to distinguish between blink and holding eyes closed for a certain period of time. Transitions between states are determined by the mean magnitude of the normal flow and its dominant orientation. The main drawback of the algorithm is that it fails when flow direction fluctuates, which leads to incorrect transitions between states in the deterministic finite state machine.

Torricelli et al. [18] proposed a simple frame differencing and image binarization with empirically determined threshold. This approach is based on the assumption, that if blink occurs then two regular blobs are obtained in a binarized image. Next, this binary image is denoised and eye's proportion is analyzed according to anthropometric knowledge.

Wu et al. [20] proposed a framework, which can be used to detect the driver's drowsiness by analysis of the eye state. In their approach, the face is localized using the Adaboost algorithm, the left eye position is detected using Support Vector Machines (SVM) classifier and features are calculated using the radial-symmetry and distance transform. The eye state is also determined by SVM classifier, but with Local Binary Patterns (LBP) features. In two experiments some video sequences from ZJU Eyeblink database was used as training set and the rest of them as test set.

In this paper we introduce some modifications to the WGT method and we present an efficient and fully-automated zero crossing detection algorithm for detection of blinks from the acquired waveforms. This work is organized as follows. The next Section provides a short description of our WGT method. In Section 3 we

present zero crossing detection technique. In Section 4 we show some modifications of the WGT method and discuss the obtained detection results. Section 5 concludes the paper.

2 The Weighted Gradient Descriptor

Our approach is motivated by the work of Polikovsky et al. [14] in which the authors have shown that a 3D gradient descriptor is capable to recognize small facial movements recorded by a high-speed camera. However, the proposed descriptor is not intuitive and difficult to analyze. In our recent paper [16], we simplified their approach and proposed a new, fast and accurate blink detection algorithm. Our basic idea was to reduce the dimensionality of the features used in [14].

The first step of the proposed scheme is the face localization using the Adaboost implementation proposed by Viola et al. [19] and modified by Lienhart et al. [10], followed by eye localization using the cascade-models described by Santana et al. in [4]. As cascade-models may return more than one eye localization and some of them can be false positive, therefore it is necessary to use other eye detection methods for validation. Hence, we changed our eye localization scheme. At the beginning we mark the right eye region manually in the first frame and then we track the eye position using the cascade-models described by Santana [4] in the small neighborhood of the eye region localized in the previous frame. We also assumed, that if the eye region position during the tracking step was not detected, then we copy the eye localization from the last frame. This simplified approach allows for correct eye localization in all of the tested video sequences.

Subsequently, we calculate the partial derivatives into y coordinate and time (t) direction for each pixel within the localized eye region for two consecutive frames of a video sequence defined as:

$$\begin{aligned} I_y(x, y, t) &= I(x, y + 1, t) - I(x, y - 1, t), \\ I_t(x, y, t) &= I(x, y, t + 1) - I(x, y, t - 1). \end{aligned} \quad (1)$$

The couples of spatio-temporal derivatives can be treated as a vector denoted as $\mathbf{I}_{yt}(x, y, t) = [I_y(x, y, t), I_t(x, y, t)]$, which represents vertical motion between frames. In this way, all vectors are divided into two groups. The first group contains all vectors for which I_t component is greater than zero and the second group consists of all vectors for which I_t is negative. Subsequently, we calculate a linear combination of the vectors with weighting coefficients equal to absolute value of the components I_t in each group. Finally, we obtain two weighted vectors $\mathbf{v}(t)_\uparrow, \mathbf{v}(t)_\downarrow$ and calculate the difference between the y -coordinates of those vectors, which will be called the waveform and denoted as $D(t)$. The weighted gradient vector $\mathbf{v}(t)_\uparrow$ is joining the initial spatial position $(v_{x0}(t)_\uparrow, v_{y0}(t)_\uparrow)$ defined as:

$$v_{x0}(t)_{\uparrow} = \frac{\sum_{x,y, I_t(x,y,t) > 0} |I_t(x,y,t)| \cdot x}{\sum_{x,y, I_t(x,y,t) > 0} |I_t(x,y,t)|}, \quad v_{y0}(t)_{\uparrow} = \frac{\sum_{x,y, I_t(x,y,t) > 0} |I_t(x,y,t)| \cdot y}{\sum_{x,y, I_t(x,y,t) > 0} |I_t(x,y,t)|}, \quad (2)$$

with a terminal point $(v_{x1}(t)_{\uparrow}, v_{y1}(t)_{\uparrow})$ defined as:

$$v_{x1}(t)_{\uparrow} = \frac{\sum_{x,y, I_t(x,y,t) > 0} |I_t(x,y,t)| \cdot (x + \delta_x \cdot I_y(x,y,t))}{\sum_{x,y, I_t(x,y,t) > 0} |I_t(x,y,t)|},$$

$$v_{y1}(t)_{\uparrow} = \frac{\sum_{x,y, I_t(x,y,t) > 0} |I_t(x,y,t)| \cdot (y - \delta_y \cdot I_t(x,y,t))}{\sum_{x,y, I_t(x,y,t) > 0} |I_t(x,y,t)|}, \quad (3)$$

where δ_x, δ_y are parameters connected with the scale of the vectors. The weighted gradient vector $\mathbf{v}(t)_{\downarrow}$ is defined analogously. The visualization of the weighted gradient vectors is presented in Fig. 1.

In the next step we compute the $D(t)$ waveform defined as:

$$D(t) = d(t) \cdot (||\mathbf{v}(t)_{\uparrow}|| + ||\mathbf{v}(t)_{\downarrow}||), \text{ where } d(t) = v_{y0}(t)_{\downarrow} - v_{y0}(t)_{\uparrow}. \quad (4)$$

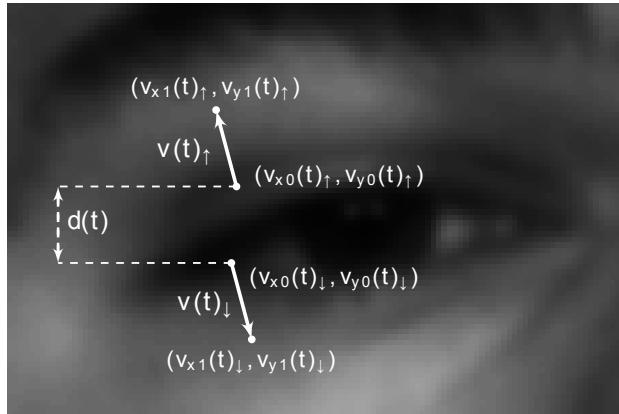


Fig. 1 The visualization of the Weighted Gradient Descriptor

The obtained waveform is noisy, therefore we propose some modifications to overcome the problems connected with accurate blink detection. More details and properties about this descriptor can be found in [16].

3 The modifications of the WTG

During the experiments we observed that it is beneficial to modify the formula describing the waveform $D(t)$:

$$\hat{D}(t) = d(t) \cdot ||\mathbf{v}(t)_\uparrow|| \cdot ||\mathbf{v}(t)_\downarrow||, \text{ where } d(t) = v_{y0}(t)_\downarrow - v_{y0}(t)_\uparrow. \quad (5)$$

This simple modification allows for reduction of noise in the waveform and facilitates easier analysis and interpretation of this curve.

Let us consider only $I_t(x, y, t)$ for each pixel and its sign. Then, we divide all pixels into two groups. The first group contains pixels for which $I_t(x, y, t) > \varepsilon$ and the second group consists of pixels for which $I_t(x, y, t) < -\varepsilon$. This assumption is necessary, because the eyelid movement causes a strong change of $I_t(x, y, t)$ values. The pixels whose $I_t(x, y, t)$ are in the range $[-\varepsilon, \varepsilon]$ are treated as noise. The center of gravity of those groups of pixels is defined as:

$$v_{x0}(t)_\uparrow = \frac{\sum_{x, y, I_t(x, y, t) > \varepsilon} |I_t(x, y, t)| \cdot x}{\sum_{x, y, I_t(x, y, t) > \varepsilon} |I_t(x, y, t)|}, v_{y0}(t)_\uparrow = \frac{\sum_{x, y, I_t(x, y, t) > \varepsilon} |I_t(x, y, t)| \cdot y}{\sum_{x, y, I_t(x, y, t) > \varepsilon} |I_t(x, y, t)|}. \quad (6)$$

The $v_{x0}(t)_\downarrow, v_{y0}(t)_\downarrow$ is defined analogously. In the next step we compute the $\tilde{D}(t)$ waveform defined as:

$$\tilde{D}(t) = d(t) \cdot \left(\sum_{x, y, I_t(x, y, t) > \varepsilon} |I_t(x, y, t)| \right) \cdot \left(\sum_{x, y, I_t(x, y, t) < -\varepsilon} |I_t(x, y, t)| \right), \quad (7)$$

where again $d(t) = v_{y0}(t)_\downarrow - v_{y0}(t)_\uparrow$. The distance between the centers of gravity of the two groups of vectors is used for the description of the eye movement dynamics.

Another problem, which should be considered is the influence of eyebrow movements and of the noise corrupting the images of the eye region. Generally, the peaks in the waveform which are caused by the eye blinks depend slightly on the position and size of the selected eye region. However, if the rectangular region of the eye is too large and the noise decreases the video quality, then the centers of gravity of the two groups of pixels can be shifted causing distortions of the waveform. This effect is especially noticeable when the eyebrow is moving. To minimize the influence of the size of the detected eye region and to diminish the influence of the eyebrow movement, we multiply $I_t(x, y, t)$ values by a two-dimensional Gaussian Kernel (GK).

The eye blink is analyzed using the waveform, in which the blink is represented by a zero-crossing point between a local maximum and a local minimum. The eye closure is exhibited by a positive peak in the waveform and the eye opening is coupled with a negative waveform peak. In the spontaneous blink the time of closing is shorter than time of opening, therefore a stronger positive peak is obtained. To denoise the waveform, we used the running average smoothing algorithm. For the detection of a blink, we propose a simple scheme:

- Calculate the maximum value D_{max} and the minimum value D_{min} in the waveform.
- Find a local maximum argument t_{max} for which $D(t_{max}) > D_{max}/n_1$ in the waveform, where n_1 is a scaling parameter and in subsequent k frames try to find a local minimum argument t_{min} for which $D(t_{min}) < D_{min}/n_2$, where n_2 is another scaling parameter and k determines the number of frames, after which the local minimum should appear. If you find a new local maximum argument t'_{max} before a local minimum, then $t_{max} = t'_{max}$.
- If you find t_{max}, t_{min} , then estimate a linear regression for the data $x = t_{max}, t_{max} + 1, \dots, t_{min}$ and $y = D(t_{max}), D(t_{max} + 1), \dots, D(t_{min})$.
- If the slope of the regression line is smaller than λ , then between a local maximum and minimum the blink is detected. The zero-crossing point is calculated as the mean of t_{max} and t_{min} .

The visualization of the detection scheme is presented in Fig. 2. The use of a linear regression is necessary, because the waveform can be noisy and simple analysis of maximum and minimum locations can provide poor results.

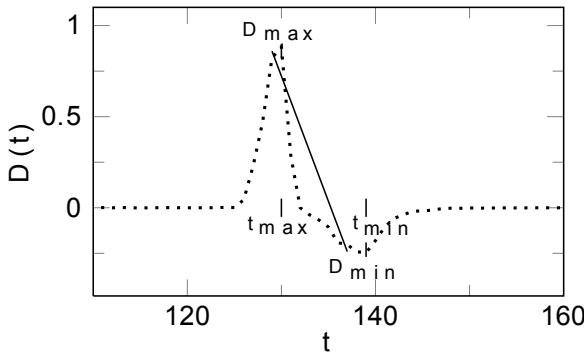


Fig. 2 The visualization for zero-crossing detection algorithm

4 Results

The proposed approach has been tested using our own database of video sequences and on the ZJU Eye blink Database described in [12]. We compare our results with other approaches validated on this database. The results will be presented using Detection Rate (DR) and number of False Positives (FP) to facilitate comparison with other blink detection techniques.

In the prepared experiment, five subjects were recorded during short interviews in a controlled environment with a Basler camera at 100 fps and 640×480 pixel resolution (one of the subjects was wearing glasses). A single video sequence contains from 8129 to 16012 frames. In five sequences we manually annotated the states of the right eye of the subjects.

Below, we present our preliminary results of analysis of the waveform for the videos recorded using a high speed camera. We used the described automatic zero-crossing detection algorithm and the results are presented in Tab. 1. As we can see, the $\tilde{D}(t)$ waveform allows to obtain relatively low false positive and a high detection rate.

Table 1 Summary of the algorithm performance (FN- Frame Number, DB - Detected Blinks, DR - Detection Date, FP - False Positive)

(a) Manually marked blinks		(b) Detected in $D(t)$ waveform		
	FN Blinks	DB	DR	FP
Person1	16013 32	31	96.88%	17
Person2	8128 33	32	96.97%	11
Person3	11498 99	88	88.89%	27
Person4	10213 55	51	92.73%	14
Person5	13179 81	80	98.77%	11

(c) Detected in $\hat{D}(t)$ waveform			(d) Detected in $\tilde{D}(t)$ waveform		
	DB	DR	DB	DR	FP
Person1	28	87.50%	28	87.50%	2
Person2	32	96.97%	32	96.97%	3
Person3	85	85.86%	87	87.88%	0
Person4	47	85.45%	39	70.91%	5
Person5	80	98.77%	77	95.06%	2

We observed a decreased efficiency of the proposed algorithm for the Person 4, which was caused by atypical blinking pattern and persistent eye movements. The differences between the original method described in [16] and the introduced modifications are shown in Fig. 1.

The ZJU Eyeblink Database contains 80 videos. The subjects were recorded in four different positions with a web camera at 30 fps and 320×240 pixel resolution. The authors claim, that the total number of blinks in the database is 255, but we observed 258 blinks. The difference can be caused in the applied definition of the blink.

The annotations were prepared using ELAN Linguistic Annotator software [2, 11]. During the tests on ZJU Eyeblink Database we obtained the best results for the \tilde{D} waveform. The used parameters are $\varepsilon = 0.02$, the standard deviation parameters in the GK were $\sigma_x = \sigma_y = 30$. We achieved 98.83% of detection rate and only 1 false positive. The achieved results are very satisfactory, however this database does not contain eye movements, eye occlusions or head movements. Some examples of the obtained waveforms are presented in Fig. 2. The results achieved by other researchers are shown below.

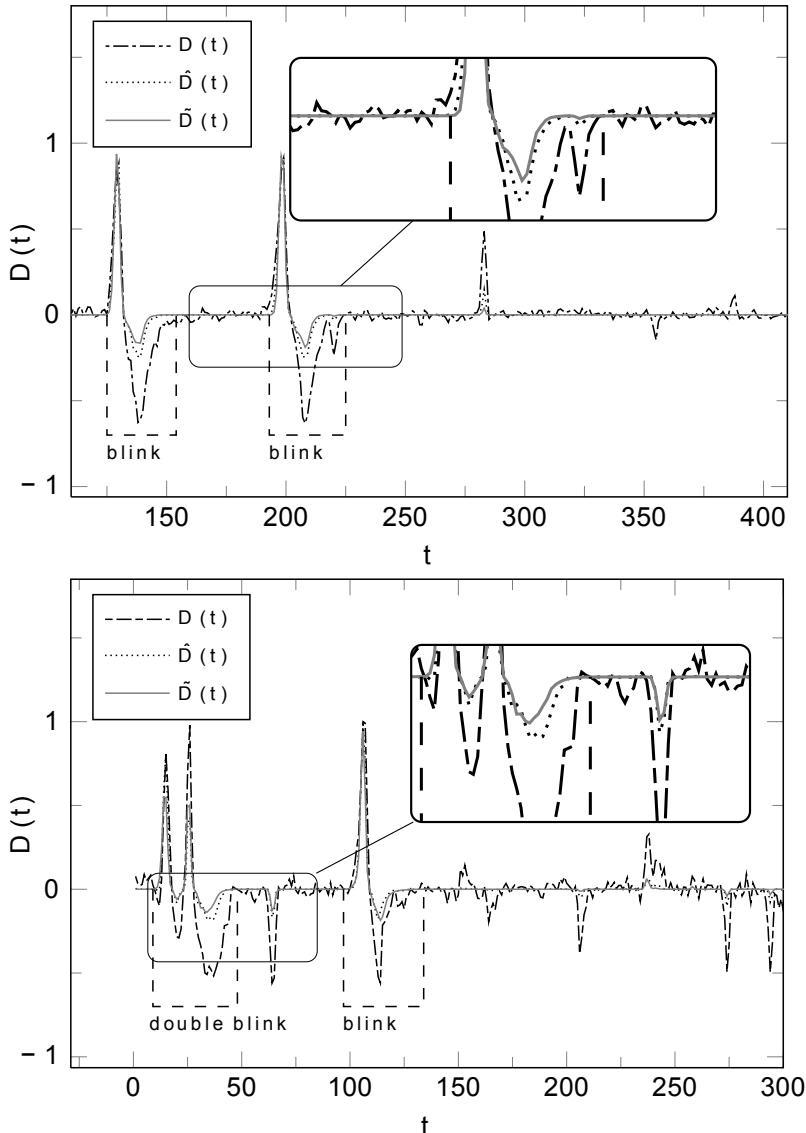


Fig. 1 The comparison of three waveforms $D(t)$, $\hat{D}(t)$, $\tilde{D}(t)$ divided by a maximal value in the waveform to show them in one plot. In the rectangular box a part of the waveform is presented in the enlarged view. The manually annotated blinks are denoted using a dashed line.

In [20] the authors presented the accuracy rate of eye state recognition. In the first experiment the accuracy was 87.13% for closed eyes and 93.62% for opened eyes. In the second experiment the accuracy was 90.11% for closed eyes and 78.44% for opened eyes. The rest of the results are shown in Tab. 2.

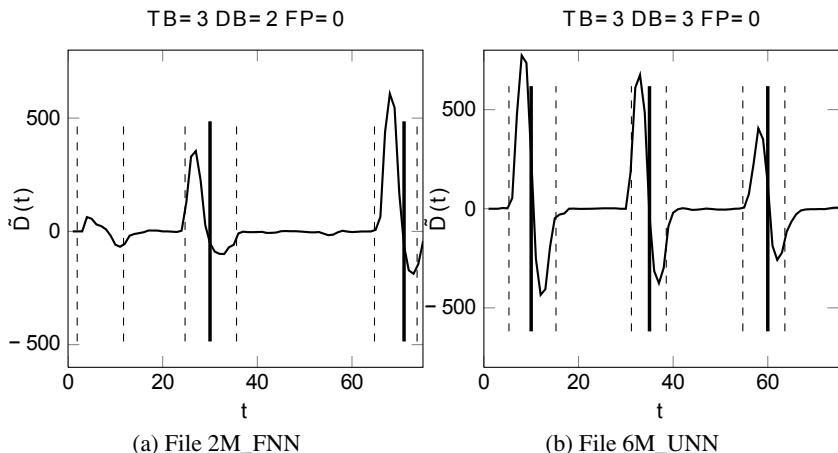


Fig. 2 Exemplary waveforms $\tilde{D}(t)$ obtained for the ZJU Blink Database. The bold lines represent the detected zero-crossings, which are signatures of the eye blinks and the dashed lines represent the points, when blinks are starting and ending.

Table 2 Comparison of the performance of eye blink detection methods tested on the ZJU Eyeblink Database

Method	FN	Blinks	DB	DR	FP
Our best results	11830	258	255	98.83%	1
Other approaches					
Conditional Random Field [7, 12]	11800	255	-	88.8% or 95.7%	<0.1%
Normal flow [6]	11800	255	-	70%	13%
Normal flow [5]	11800	255	-	97% \pm 7%	2% \pm 6%
Anthropometric properties [18]	-	-	-	95.7%	7%

5 Conclusion

In the paper, some modifications of the eye blink detection framework presented on [16] have been proposed. The novel approaches were tested on our database and ZJU Eyeblink Database. The obtained results were compared with the other methods which used the ZJU Eyeblink database. The comparison with existing techniques confirms the high efficiency of the proposed approach and its high potential for the application in systems analyzing the blinking patterns.

References

1. Abe, K., Ohi, S., Ohayama, M.: Eye-gaze Detection by Image Analysis under Natural Light. In: Jacko, J.A. (ed.) Human-Computer Interaction, Part II, HCII 2011. LNCS, vol. 6762, pp. 176–184. Springer, Heidelberg (2011)

2. Brugman, H., Russel, A.: Annotating multimedia/ multi-modal resources with ELAN. In: Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation (2004)
3. Caffier, P.P., Erdmann, U., Ullsperger, P.: Experimental evaluation of eye-blink parameters as a drowsiness measure. *European J. of Applied Physiology* 89, 319–325 (2003)
4. Castrillón, S.M., Déniz, O., Guerra, C., Hernández, M.: Encara2: Real-time detection of multiple faces at different resolutions in video streams. *J. of Visual Communication and Image Representation* 130–140 (2007)
5. Divjak, M., Bischof, H.: Eye blink based fatigue detection for prevention of computer vision syndrome. In: Proceedings of the IAPR Conference on Machine Vision Applications (MVA 2009), pp. 350–353 (2009)
6. Divjak, M., Bischof, H.: Real-time video-based eye blink analysis for detection of low blink-rate during computer use. In: First Int. Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS 2008), pp. 99–107 (2008)
7. Pan, G., Wu, Z., Sun, L.: Liveness detection for face recognition. In: Recent Advances in Face Recognition, pp. 236–252. InTech (2008)
8. Królak, A., Strumiło, P.: Eye-blink controlled human-computer interface for the disabled. In: Hippe, Z.S., Kulikowski, J.L. (eds.) *Human-Computer Systems Interaction*. AISC, vol. 60, pp. 123–133. Springer, Heidelberg (2009)
9. Leal, S., Vrij, A.: Blinking during and after lying. *J. of Nonverbal Behavior* 32, 187–194 (2008)
10. Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. In: Int. Conference on Image Processing, vol. 1, pp. I-900–I-903 (2002)
11. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands: Elan linguistic annotator software, <http://tla.mpi.nl/tools/tla-tools/elan/> (accessed: January 14, 2013)
12. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcam. In: IEEE 11th Int. Conference on Computer Vision, ICCV 2007, pp. 1–8 (2007)
13. Panning, A., Al-Hamadi, A., Michaelis, B.: A color based approach for eye blink detection in image sequences. In: IEEE Int. Conference on Signal and Image Processing Applications (ICSIPA), pp. 40–45 (2011)
14. Polikovsky, S., Kameda, Y., Ohta, Y.: Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In: 3rd International Conference on Crime Detection and Prevention (ICDP 2009), pp. 1–6 (2009)
15. Porter, S., Ten Brinke, L.: Reading between the lies: identifying concealed and falsified emotions in universal facial expressions. *Psychological Science* 19(5), 508–514 (2008)
16. Radlak, K., Smolka, B.: A novel approach to the eye movement analysis using a high speed camera. In: 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA 2012), Zouk-Mosbeh, Lebanon, pp. 139–144 (2012)
17. Stern, J.A., Boyer, D., Schroeder, D.J.: Blink rate as a measure of fatigue. U.S. Dept. of Transportation, Federal Aviation Administration, Office of Aviation Medicine (1994)
18. Torricelli, D., Goffredo, M., Conforto, S., Schmid, M.: An adaptive blink detector to initialize and update a view-based remote eye gaze tracking system in a natural scenario. *Pattern Recognition Letters* 30(12), 1144–1150 (2009)
19. Viola, P., Jones, M.J.: Robust real-time face detection. *Int. J. Comput. Vision* 57, 137–154 (2004)
20. Wu, Y.S., Lee, T.W., Wu, Q.Z., Liu, H.S.: An eye state recognition method for drowsiness detection. In: IEEE 71st Vehicular Technology Conference (VTC 2010), pp. 1–5 (2010)

Touchless Input Interface for Disabled

Adam Nowosielski and Łukasz Chodyła

Abstract. New interfaces for human-computer interaction are based on sophisticated methods of image processing and pattern recognition. They are based on some kind of visual system and offer touchless and convenient medium for communication for the user. On the other hand, there is a great number of people who suffer from some form of disability. They are unable to use standard input devices (mouse and keyboard) or take advantage from new sensors (e.g. Kinect). For people unable to control hand or body movements special assistive devices and interfaces are proposed. An interface for controlling an operating system is proposed in the article. It is based on processing video stream from a simple webcam. The solution is open and extensible. It can be used by disabled users or in hands-busy applications.

1 Introduction

According to *World report on disability* [15] over 1 billion people in the world have some form of disability. Almost one fifth of them encounter significant difficulties in their daily lives. To improve their participation in social activities and increase their independence new assistive devices and technologies are being introduced. That also includes communication and information technologies for enabling or improving services in digital world to people with disabilities. Information access is extremely important nowadays and control/input devices play the key role.

In last decades much work has been done in the human-computer interaction. Contemporary research on new interfaces focuses on computer vision and pattern recognition approaches. Hands-free control of electronic devices by natural user interface using gestures is gaining more and more attention. What seems to be natural

Adam Nowosielski · Łukasz Chodyła

West Pomeranian University of Technology in Szczecin, Faculty of Computer Science and Information Technology, Zolnierska 52, 71-210 Szczecin, Poland

e-mail: anowosielski@wi.zut.edu.pl, chodak166@op.pl

for healthy people is of no use for disabled. To increase their independence specific assistive technologies are introduced. The following techniques (addressed later in the article) are used most frequently:

- eyetracking,
- brain computer interfaces,
- speech recognition,
- computer vision approaches.

The rest of the article is structured as follows. In Sect. 2 touchless interfaces for disabled are considered. Section 3 describes the proposed solution and some details are provided in Sect. 4. The article ends with a summary.

2 Touchless Techniques for Input Interfaces

There are four main techniques for touchless interfaces. The first one - eyetracking - offers sophisticated methods for capturing humans' gaze direction. Eye movements are used as control medium for human computer interaction (HCI) or for activity analysis (usually off-line systems) [5]. In the first case eye movements replace traditional control devices or are used supplementarily by affecting the interface in real time. Examples include supporting multimedia learning [2] or graphics systems [7]. In case of off-line systems eye movements do not affect the user interface but are recorded and used in: psychology and vision research, cognitive studies, usability analysis, interface analysis, user experience testing (e.g. web page viewing behaviour), market research, etc. The main reason for low popularity is the high cost of commercial eye trackers. However, papers on low-cost eyetrackers construction appear in the literature [8].

The most common eye trackers capture image of the pupil by a video camera sensitive to the infra-red spectrum. The eyes are exposed to direct invisible infra-red (IR) light and corresponding corneal reflections are measured enabling the estimation of observer's gaze point [8].

An example of proprietary solution is the CEye system from Tobii [13]. It is designed for people with Cerebral Palsy, ALS, stroke/aphasia, spinal cord injuries and other conditions. The movement of the eyes - eye control - is directly interpreted for mouse movement. To select an icon or the letter a simple blink is used.

In Brain-Computer Interfaces (BCI) presence of specific patterns in a person's brain activity is analysed. Any form of communication begins with the user's intent which activates certain brain areas. The nervous system transmits signals to the corresponding muscles. This triggers the movement process necessary for the communication or control task. The BCI bypasses this process. Measured brain activity associated with the user's intent is converted into control signals for BCI applications [3]. Methods of processing the brain signals seems to play the vital role here [1]. BCI interfaces attract great attention with new proposals and prototypes being introduced.

Speech is another medium for human-computer interaction. Speech recognition technology has been a topic of research since the 1950s [11]. It has great potential in dialogue systems and in device control (smartphones, operating systems). Furthermore, speech is used for dictation eliminating the need to type. This technology is mature and for those who are able to use it, should be considered as a significant medium in HCI.

Finally, to provide people with disabilities the way to manipulate in electronic world, computer vision techniques are used. There are solutions operating in visible and infrared light. Most approaches use head or marker tracking. In the first case control over a computer is allowed by movement of the head. First, the user face is detected and then tracked by one of the many methods [17]. An example of mouse replacement with head tracking is the Enable Viacam (eViacam) [6]. It is markerless solution that works in visible spectrum. It works on standard PC equipped with a webcam. Face and mouth detection/tracking are of most importance here.

In marker systems special visible distinctive material is used for easy foreground separation. Markers are usually placed on the selected part of the body. SmartNav from NaturalPoint [9] is an example complete product. It works in infrared light and uses: infrared emitters, reflector and an infrared camera. It offers hands-free cursor control with movements of the head or any other part of the body.

Many computer vision approaches take advantage from the user eyes. These solutions, however, can not be considered as eyetracking techniques since no user gaze point is tracked. The image plane position of the eyes is detected and its localization changes are transferred to a control system [12]. Eye blink is frequently used for confirmations. Systems which utilize this technique must discriminate between two blinking types, namely involuntary blink and steering blink. They should differ in time duration. An interesting project of text entry solely with eye blink is presented in [10]. The interface consists of the screen keyboard with periodically highlighted columns and rows. The user makes a selection with the eye blink.

3 Interface Conception

The use of standard input devices for some people is limited to their disabilities. The current price level of special assistive devices is far too high for many individuals. These solutions cannot be extended independently of the proprietor. Appearance of low-cost products and open source applications seems to be of great importance here.

Presented approaches and solutions (Sect. 2) show that hands-free mouse alternative may be successfully achieved with the use of camera and body (usually head) movements. However, many of them lacks versatility. They are limited only to mouse or keyboard input. Additionally, they are confined how the user should interacts (detection of the single feature without the possibility of change).

Two most important tasks in human-computer interaction include text input and mouse control. An interface for controlling an operating system is proposed here as

an alternative for a physical device. The main elements include the onscreen keyboard and the visual gesture interpreter. The main interaction is through the movement of the user, recognized by processing video stream from a simple webcam. User movements, interpreted as gestures, enable a key on the keyboard selection or cursor control in the mouse mode. The user have the possibility to link a gesture to an external system procedure (e.g. to launch an application: web browser, mail client, text editor, etc.). The gesture is defined as an ordered set of moves (referred further as actions). Recognized gesture triggers the defined procedure. Actions composed from horizontal and vertical movement are only accepted.

Figure 1 shows the QVirtboard. It is the final interface displayed with two variants of feedback presentation to the user. The first mode (denoted by 'A' letter) presents cropped image acquired by the camera with guiding lines and rectangles. In second mode ('B') an action map is provided. Both modes will be addressed later. They can be set interchangeably by the user. For the experienced, the feedback might also be turned off. For those who begin - it is indispensable.



Fig. 1 QVirtboard - the onscreen interface

It was assumed that presented non-contact control interface should be open and extensible for new steering mechanisms. For this purpose, the Linux operating system was selected as the target platform and appropriate programming structures for extensions were adopted. The details, however, are omitted in this study.

4 The Interface Principles and the Gesture Detection

To use computers with the proposed interface a webcam and the presence of single or multiple markers are needed. The user is assumed to possess a characteristic, distinctive colour feature. It is used as a marker. It should be of solid colour and examples include: part of the clothing, colour eyeglasses, body part, head (face), mouth, head cover (a cap, a band or a hat), microphone boom etc. The main idea is to use any available object provided it is sufficiently distinctive in the individual user surroundings. It was assumed that user should work without additional special hardware equipment (e.g. infrared lighting, infrared camera, additional sensors, etc.).

The most important part here is the calibration process. The user can select single or multiple markers (which will be addressed later). The procedure base on the single point selection. Pixels from the closest neighbourhood are verified for similarity. Those similar are used for the generation of the colour marker model. The HSV colour space was chosen. The HSV colour space separates the luminance and the chrominance signals and makes it possible to correctly process the colour information. Examples of image processing using the colour information include face detection [14] [4], facial features localisation [16] [4] and many others. Four markers and their corresponding image masks are presented in Fig. 2 (white denotes pixels corresponding to the colour marker model).

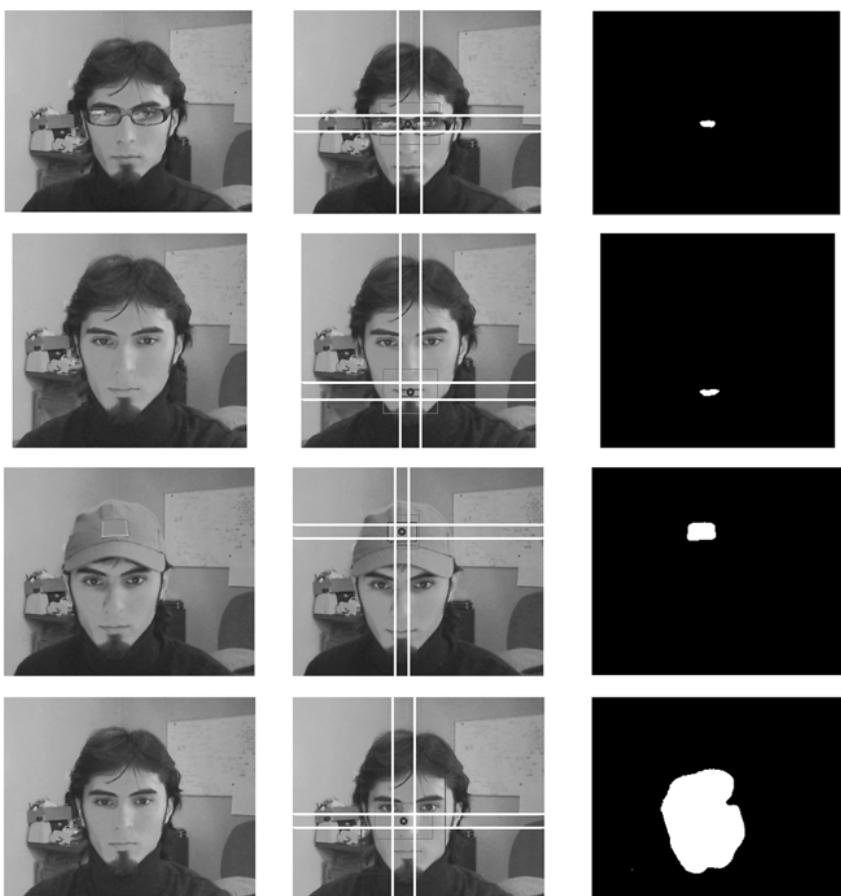


Fig. 2 Example markers (left column), detection result (middle column) and image masks (right column). Types of markers from above: a sticker, mouth, part of a cap, the whole face.

Beside colour, some background modelling mechanisms are incorporated into the method of marker detection. Since it is the user who moves in front of the camera the background detection and adaptation excludes the moving objects. When the user is still, no movement is detected. While he/she moves some movement is detected and used additionally in marker detection process.

4.1 Gestures and Actions

The idea of gesture interpretation relies on marker projection on the action map (marker plane). The action map allows proper classification of the user movement. The idea is presented in the scheme (A) on the Fig. 3 and in the form of the final appearance of the marker field (B). Neighbourhood of the detected marker is surrounded by certain neutral area - the gray area on the figure (A). This prevents accidental actions induced by noise or subtle and natural position changes.

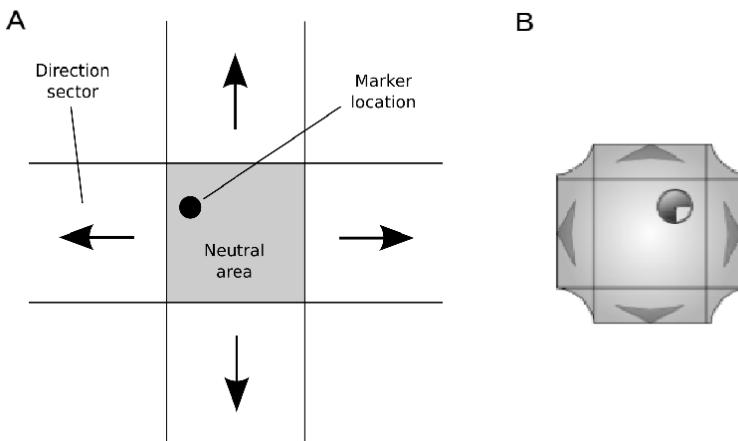


Fig. 3 Action map: the scheme (A) and the final appearance (B)

The marker movement from the neutral position, so called the dead zone, to a direction sector is interpreted as an action an added to the corresponding movement sequence. The gesture thus is defined as an ordered set of moves (actions). Recognized gesture triggers the defined procedure. The selection of the key on the onscreen keyboard is an example. However, it should be the user or his attendant, who configures gestures and corresponding reactions.

The problem with the action map is with the change in the starting position. It appears when user moves and adopts a new position in front of the camera. The possibility of dynamic and automatic correction should be provided. Without this additional mechanism many similar actions would be registered. Hence the action is only detected when a shift from the dead zone to a direction sector is followed by a

return to the dead zone. Otherwise, new starting positions are adopted. Duration time is crucial here. Experiments conducted with volunteers showed that at the beginning users tend to extend the default time but after some experience they go back to the original settings.

To help the user with the control over the time, marker was designed in a form of animated circle shape. The area of the pointer field is gradually decreased. When it takes a 360 degree turn inside a direction sector its current position will be updated as new central position. This also prevents from subtle position changes by the user inside the dead zone. Only significant changes are interpreted as actions. Naturally, user movement ability is taken into consideration hence the interface is fully adjustable.

4.2 *Marker Modes*

There are two marker modes: the tilt mode and the rotation mode. In the first mode, movement of the single marker is tracked inside the image frame. Its relative position is used for the control process. With the rotation mode two markers placed on the edges of a rotated object are needed. Their movement around the vertical axis is tracked. Their captured positions in the image plane are not taken into account. Their size and relative position to one another are relevant. For example, two markers placed at the edges of the glasses will change their mutual location during head rotation. The first marker's visibility will be limited while the second extended. This would allow the gestures interpretation independent from the location in the image plane. This will also reduce the number of possible actions by half.

5 Conclusions

There is a great number of people with disabilities who cannot have an effective control over electronic devices. Some control over computer equipment is a way to enable them communication and information access.

In the article a visual input interface was proposed. It was constructed in a form of onscreen keyboard with the capacity of mouse control. The gesture interpreter processes a video stream from a simple webcam. The user interaction is through movement (usually head). User movements interpreted as gestures allow a key on the keyboard selection or mouse operation. Such solution proved to work stable and efficiently.

It was assumed that presented non-contact control interface should be open and extensible for new steering mechanisms. The project is accessible through the sourceforge.net service (<http://sourceforge.net/projects/qvirtboard/>) and the homepage (<http://qvirtboard.sourceforge.net/>). The aforementioned homepage includes video material with examples of system operation.

The main drawback of proposed approach is the need of configuration with traditional input devices. The most challenging task here seems to be the development of a method for automatic configuration. Some steps have been already taken, for example with automatic marker detection with a simple mouse click during the calibration process. The use of head movements (tracked by a face detection algorithm) by default with a new user seemed to be a good solution. It would allow further interface modifications and adjustments. Unfortunately, this proved to be cumbersome for the users since natural head movements are relatively big causing the steering movements to be comparatively greater.

It must be noticed at last, that compared to traditional input devices (mouse and keyboard) the proposed interface is much slower. Nevertheless, for some people that might be the only way to interact with computers.

References

1. Bashashati, A., Fatourechi, M., Ward, R.K., Birch, G.E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering* 4(2), R35–R57 (2007)
2. Gog, T., Scheiter, K.: Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instructions* 20(2), 95–99 (2010)
3. Graimann, B., Allison, B.Z., Pfurtscheller, G.: Brain-Computer Interfaces: a gentle introduction. In: Graimann, B., Allison, B.Z., Pfurtscheller, G. (eds.) *Brain Computer Interfaces Revolutionizing Human Computer Interaction*. Springer Publishing, New York (2010)
4. Hsu, R.-L., Abdel-Mottaleb, M., Jain, A.K.: Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(5), 696–706 (2002)
5. Jacob, R.J.K., Karn, K.S.: Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In: *The Minds Eye: Cognitive and Applied Aspects of Eye Movement Research*. Elsevier Science, Oxford (2003)
6. Loba, C.M.: Enable Viacam (project page) (2012),
<http://eviacam.sourceforge.net/index.php>
(accessed January 19, 2013)
7. Mantiuk, R., Bazyluk, B., Tomaszewska, A.: Gaze-Dependent Depth-of-Field Effect Rendering in Virtual Environments. In: Ma, M., Fradinho Oliveira, M., Madeiras Pereira, J. (eds.) *SGDA 2011. LNCS*, vol. 6944, pp. 1–12. Springer, Heidelberg (2011)
8. Mantiuk, R., Kowalik, M., Nowosielski, A., Bazyluk, B.: Do-It-Yourself Eye Tracker: Low-Cost Pupil-Based Eye Tracker for Computer Graphics Applications. In: Schoeffmann, K., Merialdo, B., Hauptmann, A.G., Ngo, C.-W., Andreopoulos, Y., Breiteneder, C. (eds.) *MMM 2012. LNCS*, vol. 7131, pp. 115–125. Springer, Heidelberg (2012)
9. NaturalPoint, SmartNav (product page) (2009),
<http://www.naturalpoint.com/smarnav/>
10. Orange Labs Poland, Technical University of Lodz, b-Link (project page) (2009),
<http://b-link.sourceforge.net/> (accessed January 19, 2013)
11. Rebman, C.M., Aiken, M.W., Cegielski, C.G.: Speech recognition in the human-computer interface. *Information & Management* 40(6), 509–519 (2003)
12. Santis, A., Iacoviello, D.: Robust real time eye tracking for computer interface for disabled people. *Computer Methods and Programs in Biomedicine* 96(1), 1–11 (2009)

13. Tobii Technology, CEye (product page) (2012), <http://www.tobii.com/en/assistive-technology/global/products/hardware/ceye-eye-control-module/> (accessed January 01, 2013)
14. Vezhnevets, V., Sazonov, V., Andreeva, A.: A Survey on Pixel-Based Skin Color Detection Techniques. In: Proc. GRAPHICON 2003, pp. 85–92 (2003)
15. World Health Organization, The World Bank, World report on disability. WHO Press, World Health Organization, Switzerland (2011)
16. Xia, Y., Mao, K., Wang, W.: A Novel Facial Feature Localization and Extraction Method. *Journal of Computational Information Systems* 6(8), 2779–2786 (2010)
17. Yang, M.-H., Kriegman, D.J., Ahuja, N.: Detecting Faces in Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(1), 34–58 (2002)

Architecture of the Semantically Enhanced Intellectual Property Protection System

Dariusz Ceglarek

Abstract. This paper presents the structure and the functionality of the Semantically Enhanced Intellectual Property Protection System. The system uses an extensive set of semantic net algorithms for the Polish and English language that which allows it to detect similarities between compared documents on a level far beyond simple text matching. SeiPro2S benefits result both from using a local document repository and from Web based resources. The SeiPro2S system uses a mechanism of semantic compression developed to generalize concepts during a comparison of documents. The main focus of this work is to give the reader an overview of architecture, applied mechanisms and some actual results.

1 Introduction

The SeiPro2S system architecture was first described in [4]. The system has been designed to protect resources from the unauthorized use of intellectual property, including its appropriation. The most commonly known example of such treatment is plagiarism, which, as is well known, infringes intellectual property by hiding the origins (e.g. reusing someone else work without changes to the narration, examples and the order of arguments). The system is able to detect plagiarism and also is able to perform tasks such as the protection of intellectual property in documents assigned to the system (monitoring mode) in which the system periodically monitors documents listed in the available repositories of documents (e.g. the Internet) and detects documents violate the intellectual property. In such cases a list of permissible locations of document copies is prepared for a protected document.

Classic methods employed by plagiarism detection systems base on simple knowledge representation, in which a document is a set of words that were used.

Dariusz Ceglarek
Poznan School of Banking, Poland
e-mail: dariusz.ceglarek@wsb.poznan.pl

This representation yields well-established methods of classification using for example vector space model [1]. The SeiPro2S system uses semantic network as a structure responsible for knowledge representation. Its greatest advantage is by supplying a system with the right meaning of the concept processed based on its contextual usage. Benefits one can obtain by applying semantic nets in classification tasks were described by [1]. For processing documents written in Polish a semantic net SenecaNet was employed in the SeiPro2S system. In [6] have been shown that the automated method of the SenecaNet network expansion with new concepts and new lexical relationships using specially constructed transducer enabling that SenecaNet network has comparable size and performance quality as WordNet realizing Natural Language Processing (NLP) tasks.

The SeiPro2S uses a mechanism of semantic compression developed to generalize concepts, allowing to analyze the similarity of the content of compared documents using similar concepts (expressions having similar semantic meaning) to the base document. The system can be used (research mode) based on a detailed description of the research problem, it allows to automatically build a relatively complete reference base for some research domain. This would allow the finding of documents dealing with the same idea, but expressed in a different way. Task realized by SeiPro2S for English has been made possible through adoption of Wordnet and adjusting it to already existing tools crafted for the SEIPro2S system. The details of the adoption and motivation of transferring Wordnet to SenecaNet format is discussed in [5].

For the task of finding sources for a suspect document (searching for plagiarism) working prototype allows to produce a report demonstrating areas copied or paraphrased from other sources along with clear indicator informing user on ratio of copied work to the overall document content.

With the developed algorithms using semantic networks developed for the Polish language as well as English and semantic compression as a mechanism of generalization of concepts, SeiPro2s system is able to detect the copied passages, even if the author was trying to hide his attempts by changing the word order, has been using synonyms, troponyms, hyponyms or commonly interchangeable phrases when compiling his document.

The structure of the paper is as follows: there is a brief discussion of the technologies used which is followed by a description of the system functioning, its architecture and performance, with examples from the experiments undertaken and finally a summary along with an indication of future research directions.

2 SeiPro2S Functionality and Architecture

A sequence of steps which has to be undertake when a document is to be subjected to similarity check against local document repository and documents obtained from the Internet.

When the input document is presented to the system, responsible procedure starts of sampling Internet using fragments obtained from input document. A list of

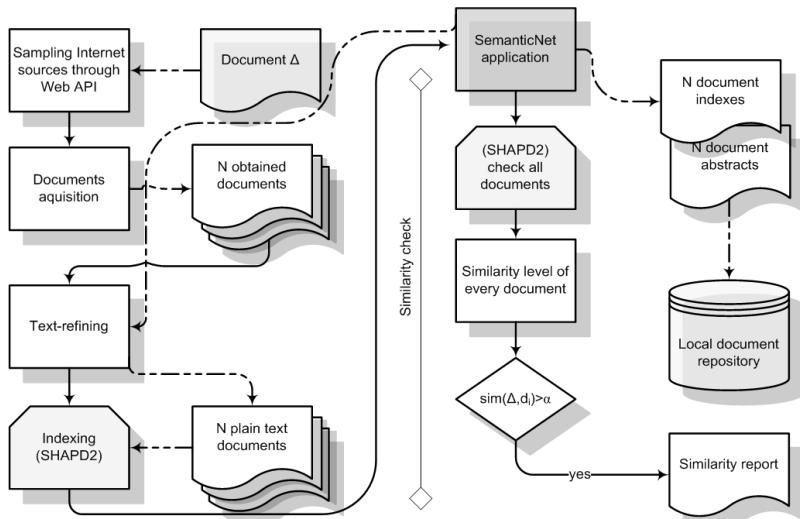


Fig. 1 Checking document against a sample of documents obtained from open Internet

potentially viable documents is prepared basing on the occurrence of necessarily long identical passages. Documents are downloaded and subjected to text-refining. After completing these steps every downloaded document is indexed. For every document an abstract is created and stored in local repository. A key for the whole process procedure follows. At first, the most relevant of previously sampled, abstracted and indexed documents are selected by comparing the indexes. Then, index of input document is subjected to comparison with hash table containing all keys coming from local corpora of documents. As a result of this action, a similarity report is constructed. It conveys following information on the overall similarity of input document to every checked document:

- length of the longest identical phrase obtained from the documents from the sampling phase
- similarity indicator, which using percent ratio demonstrates how much of the submitted text is identical with documents coming both from the sampling phase and from local repository coming from earlier checks
- checked text along with markup showing which text fragments are identical to those coming from local repository and current Internet sample.

When SEIPRO2S acts in monitoring mode some enhancements are introduced. First of all, each analyzed document along with first batch of samples obtained from the Internet is stored in the local repository. Thus, when the system is monitoring the Internet it can omit all previously checked documents and focus only on the new ones. Figure 1 presents a sequence of steps which are undertaken when a document is to be subjected to similarity check against local document repository and documents obtained from the Internet.

2.1 System Architecture

Components constituting SEIPro2S system are illustrated on Figure 2. Its main part is a semantic net module, which works as system engine processing semantic relations. The end user communicates with the system through a Web interface by defining tasks to be completed by the system. After task is done, the user can view a report. Other components of SEIPro2S are the sampling agent, downloading agent, file format adapters, abstractor, classifier and document corpus - as described below.

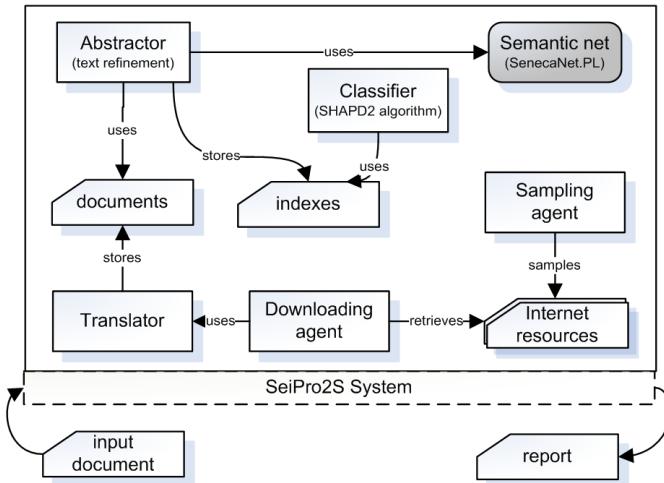


Fig. 2 SEIPro2S componentents

The sampling agent uses Internet search engines to find documents relevant to examined text. It uses document abstracts and a semantic network module to generate multiple queries. Thanks to heuristics, the agent makes use of domain dictionary to narrow down obtained results to specified domain by determining a frequency of domain-typical words. It can also query the search engine with long text phrases. The sampling agent outputs a list of URLs and passes it to the downloading agent, which downloads relevant documents from the Internet and saves them to local repository. For efficiency matters, the downloading agent works in multiple threads. Sometimes is necessary to homogenize format of retrieved resources. The file format adapter follows a well tested approach of flattening any structured file formats into plain text.

The effect of text-refinement procedure, as performed by the abstractor module, are the abstracts of documents on which the index is generated (the process of indexing the documents is described in section 2.4).

The SEIPro2S system implements innovative algorithms which ensure, that document comparison methods are insensitive to techniques used to disguise plagiarism, like changes in word order or using synonyms. A classifier module using

SHAPD2 algorithm compares documents and searches common phrases which are long enough (phrase length is a configurable parameter). The classifier uses document indexes containing hashing keys for whole sentences.

One more feature of the SEIPRO2S classifier is taking into account cases, when text copying is allowed (e.g. citations marked with quotation marks and references).

The system applies many of well known text mining techniques employing a variety of algorithms coming from various branches of computer science.

2.2 *Text Refinement*

Each task performed in the SeiPro2S needs to carry out text-refinement process (from unstructured text document input to a structure containing stacked sequentially descriptors of concepts found in the input document). Action that make up the process of text-refinement in documents starts from extracting lexical units (tokenization), and further text refinement operations are: elimination of the words from the so-called information stop-list, the identification of multiword concepts, bringing concepts to the main form by lemmatization or stemming. It is particularly difficult task for highly flexible languages, such as Polish or French (multiple noun declination forms and verb conjugation forms).

Synonyms need to be represented with concept descriptors using semantic network. It allows correct similarity analysis and also increases classification algorithms efficiency without loss in comparison quality [15].

Abstracting process faces another problem here, which is polysemy. One word can represent multiple meanings, so the apparent similarity need to be eliminated. It is done by concept disambiguation, which identifies word meaning depending on its context, is important to ensure that no irrelevant documents will be returned in response to a query [16]. A very efficient method of concept disambiguation has been proposed in [3]. It uses semantic network for Polish language and examines word context to determine its meaning, resulting in 82% accuracy.

The last operation in text refinement procedure in SeiPro2S is a generalization of concepts using semantic compression (see section 2.3 for details).

The final effect of refinement procedure is the structure of documents containing ordered descriptors of concepts derived from the input document. This structure can be stored as an abstract (data for creating index) of the document, and then use the appropriate algorithm for the implementation of specific classification task, which may be searching for long common sequences in documents in order to determine the extent of the originality of the analyzed document.

To build indexes in the SeiPro2S - both in plagiarism detection task as well as in the task of monitoring violations of intellectual property contained in the documents entrusted to the system - the system uses a special structure based on hashed whole sentences, which greatly accelerates the process of joint detection of long passages (detailed in section 2.4).

Table 1 Example of semantic compression

Sentence A	The life span of a cell depends upon the wear and tear on that cell
Sentence B	Cells lifetime reposes on accumulated damage
Sentence C	Our treasured color TV set and refrigerator was traded so that our household could carry on through
Sentence A after compression	period time cell rely damage cell
Sentence B after compression	cell period time rely accumulate damage
Sentence C after compression	our love colour television fridge sold family survive

2.3 Semantic Compression

Semantic compression can be defined as an effective technique to generalize concepts, which adapts to the context and take into account the additional requirement of minimizing the loss of information. The idea of semantic compression was introduced in 2010 in [7] as a method of enhancing the quality and efficiency of classification of text documents. Text compression is possible through the use of semantic net and frequency dictionaries. Concepts with the lowest number of occurrences are considered unnecessary and they are replaced by their synonyms and their generalizations (hypernyms stored in the semantic net). The aim of this procedure is to reduce the number of concepts used to represent concepts that occur in text documents without significant loss of information, which is very important from the perspective of natural language processing [1]).

In addition, reducing the number of concepts helps in coping with problematic linguistic phenomena which usually draws attention in NLP: polysemy and synonymy [19] or [16]. When multiple terms used to describe the same or very similar concept occur relatively rarely, they can be replaced by one common, more general, concept. Due to employment of statistical analysis in big corpora of documents, remarked frequency dictionaries are prepared and let system deal with polysemic words with less effort and lower error rate than solutions not employing this technique.

In general, the semantic compression enables information retrieval tasks, such as text matching, to operate on a concept level, rather than on level of individual terms. This can be achieved not only by gathering terms around their common meanings (known from synset based approach [18]), but also replacing longer phrases with their more compact forms.

The emphasized concept level allows for capturing of common meaning expressed with different set of terms. The process of concepts generalization in semantic compression mechanism is depicted in Figure 3. As a descriptor list for total N concepts there are selected M concepts with top values of cumulated frequency. A very simple example given in Table 1 can be used to demonstrate the idea of semantic compression.

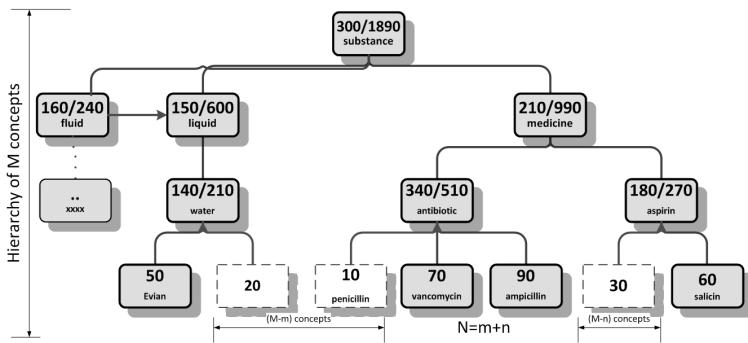


Fig. 3 Selection of M concepts with top cumulative frequencies

2.4 Searching for Common Sequences (SHAPD2)

The task of matching a longest common subsequence is the most important one in pliarism detection task. Introduction of special structure based on hashing technique that was later referenced to as a shingling (a continuous sequence of tokens in a document) can substantially improve the efficiency of deciding on the level of similarity of two documents by observing a number of common shingles [17]. A number of works represented by publications such as [13] or [2] provided plausible methods to further boost measuring of the similarity between entities. Algorithms based on shingle technique are commonly utilized to identify duplicates or near-duplicates because of their ability to perform clustering tasks in linear computational complexity [14].

The SHAPD2 algorithm (described in [10]) allows for a robust and a resilient computation of a longest common subsequence shared by one or many input documents. SHAPD2 is improved and much faster version of the SHAPD algorithm, which was described in [8].

The SHAPD2 focuses on whole sentence sequences. A natural way of splitting a text document is to divide it into sentences and it can be assumed that documents containing the same sequences, contain the same sentences. However, in text documents, there are situations in which there are long passages of text without full-stop mark (such as different types of enumerations, tables, listings, etc.). Some sort of strategy needs to be devised for such cases, i.e. how to split portions of text, which are longer than reasonable length sentence in a natural language. SHAPD2 utilizes a brand new mechanism to organize the hash-index as well as to search through the index.

Indexing

After text-refinement process all documents need to be split into text frames of comparable length - preferably sentences, or in case of longer sentences - shorter phrases (frames). The algorithm uses special coefficient which allows to set the expected

number of frames which a longer sentence is split into. The process is driven by a modular additive hashing¹ function. Every concept in a sentence is hashed by assigning a number from a previously defined range. Further, the individual hashes are summed to represent a sentence. Thanks to the additive nature of hashing function, sentences with changed concept order are treated as equivalents.

Thus, the resulting algorithm not only searches the longest common sequences, moreover the longest common quasi-sequences (allowing minor editing changes such as syntactic changes, insertions, deletions, and synonyms replacements, as well as combining or splitting multiple sentences to change their structure).

Detection Long Common Sequences

The SHAPD2 algorithm is able to compare two document's corpora in single pass. There are two corpora of documents as the algorithm's input: a corpus of source documents (originals) $D = \{d_1, d_2, \dots, d_n\}$, and a corpus of suspicious documents to be verified regarding to possible plagiaries, $P = \{p_1, p_2, \dots, p_r\}$.

In first step, a hash table T is being created for all documents from corpus D , where for each key the following tuple of values is stored: $T[k_{i,j}] = <i, j>$, (document number, frame number).

For phase 2, a correspondence list CL is declared, with elements of the following structure: nd - document number, ml - local maximum, and nl - frame number for local sequence match. Another data structure is maxima array TM for all r documents in corpus P , containing records with structure as follows: mg - global maximum, ng - frame number with global sequence match.

For all documents d_i from corpus P (suspicious documents), the correspondence array CL and maxima array TM are cleared. For each frame, set of tuples is retrieved from index table T . If there are any entries existing, it is then checked whether they point to the same source document and to the previous frame. If the condition is true, local correspondence maximum is increased by one. Otherwise, the local maximum is decreased. After all frames are checked, table TM storing correspondence maxima is searched for records whose correspondence maxima is greater than a threshold set φ (number of matching frames to be reported as a potential plagiarist). Frame and document number is returned in these cases.

Phase 2 is performed sequentially for all documents from corpus P .

In order to evaluate the algorithm's efficiency, a series of experiments was executed with Reuters corpus consisting of multiple sets of annotated documents (originals as well as suspicious documents). A set of 3000 original documents has been used as source set, and several sets including 1000 to 6000 documents have been used as a set suspicious documents. The sets were compared using two algorithms: w-shingling and SHAPD2. All tests have been executed on one computing platform, a stock laptop computer, clocked at 2.0 GHz.

¹ Hashing is a commonly used technique in NLP tasks. In plagiarism detection is used to identify common word sequences as well as to identify duplicates [14].

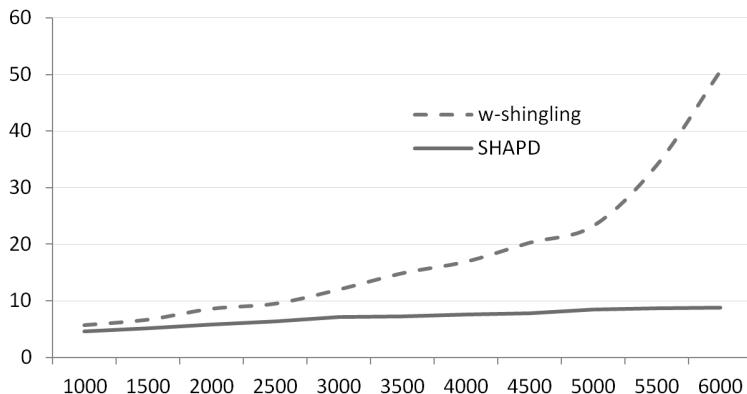


Fig. 4 Processing times (milliseconds) of comparing n documents with a corpus of 3,000 documents using *w-shingling* and *SHAPD2* algorithms

The basic results of efficiency test shows that a comparison of one suspicious document to a set of 3,000 originals (containing about 3060 words on average) takes 7.13 ms (see Figure 4). As many as 420,700 document-to-document comparisons were achieved in 1-second intervals. The resulting times of tests' execution are presented in Table 2.

Table 2 Processing time [ms] for comparing n suspicious documents with a corpus of 3,000 original documents

n	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500	6000
w-shingling	5680	6654	8581	9478	11967148641689920242232003395550586						
SHAPD2	4608	5114	5820	6374	7125	7213	7527	7818	8437	8656	8742

Moreover, the algorithm is utilizing the same techniques, that have already proven their effectiveness in plagiary detection employing semantic compression [7], as well as a resilience to false-positives examples of plagiarism [12], which is may an issue in case of using competitive algorithms. Clough & Stevenson Corpus of Plagiarized Short Answers [11] was used for the benchmark.

3 Summary

The SEIPro2S system has proved to be a efficient tool in checking whether submitted content is not an unauthorized copy. SeiPro2S makes it possible to not only find direct copying, but also to find passages that rephrase the copied content with another set of words, thus reproducing the original thought. Designed SHAPD2

algorithm is highly efficient in plagiary detection task and employing semantic compression is strong resilient to false-positives examples of plagiarism (see [9]), which is may an issue in case of using competitive algorithms.

Working as a authenticity checker is only one possible application of the SeiPro2S system. Among others, one has to emphasize it as a tool that can monitor whether important information does not leak out of organization. In addition, system can be of great help to researchers checking whether their idea was addressed by their peers granting them with better overview of chosen research field.

There are also few topics for further research. Certain phrases are typical for some domains (eg. law, science) and therefore their exact occurrence in different documents cannot be perceived as the act of plagiarism. Next goal is applying FrameNet approach for establishing equivalent phrases not detecting by semantic compression mechanism. An investigation how to recognize allowed proper citations, could also be done.

References

1. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press, Addison-Wesley Longman Publishing Co., New York (1999)
2. Broder, A.Z.: Syntactic clustering of the web. *Comput. Netw. ISDN Syst.* 29(8-13), 1157–1166 (1997)
3. Ceglarek, D.: Zastosowanie sieci semantycznej do disambiguacji pojed w jezyku naturalnym. In: Systemy wspomagania organizacji SWO 2006, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice (2006)
4. Ceglarek, D., Haniewicz, K., Rutkowski, W.: Semantically Enhanced Intellectual Property Protection System - SEIPro2S. In: Nguyen, N.T., Kowalczyk, R., Chen, S.-M. (eds.) ICCC 2009. LNCS, vol. 5796, pp. 449–459. Springer, Heidelberg (2009)
5. Ceglarek, D., Haniewicz, K., Rutkowski, W.: Quality of semantic compression in classification. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) ICCC 2010, Part I. LNCS (LNAI), vol. 6421, pp. 162–171. Springer, Heidelberg (2010)
6. Ceglarek, D., Haniewicz, K., Rutkowski, W.: Towards knowledge acquisition with wiSENNet. In: Nguyen, N.T., Trawiński, B., Jung, J.J. (eds.) New Challenges for Intelligent Information and Database Systems. SCI, vol. 351, pp. 75–84. Springer, Heidelberg (2011)
7. Ceglarek, D., Haniewicz, K., Rutkowski, W.: Semantic compression for specialised information retrieval systems. In: Nguyen, N.T., Katarzyniak, R., Chen, S.-M. (eds.) Advances in Intelligent Information and Database Systems. SCI, vol. 283, pp. 111–121. Springer, Heidelberg (2010)
8. Ceglarek, D., Haniewicz, K.: Fast plagiarism detection by sentence hashing. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2012, Part II. LNCS, vol. 7268, pp. 30–37. Springer, Heidelberg (2012)
9. Ceglarek, D., Haniewicz, K., Rutkowski, W.: Robust Plagiary Detection Using Semantic Compression Augmented SHAPD. In: Nguyen, N.-T., Hoang, K., Jędrzejowicz, P. (eds.) ICCC 2012, Part I. LNCS, vol. 7653, pp. 308–317. Springer, Heidelberg (2012)

10. Ceglarek, D.: Single-pass Corpus to Corpus Comparison by Sentence Hashing. In: Proceedings of 5th International Conference on Advanced Cognitive Technologies and Applications Conference. Xpert Publishing Services, Valencia (2013)
11. Clough, P., Stevenson, M.: A Corpus of Plagiarised Short Answers, University of Sheffield (2009), http://ir.shef.ac.uk/cloughie/resources/plagiarism_corpus.html
12. Frakes, W.B., Baeza-Yates, R.: Information Retrieval: Data Structures & Algorithms. Prentice-Hall (1992)
13. Hamid, O.A., Behzadi, B., Christoph, S., Henzinger, M.: Detecting the origin of text segments efficiently. In: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, vol. 7(3), pp. 61–70 (2009)
14. Hoad, T.C., Zobel, J.: Methods for identifying versioned and plagiarized documents. Journal of the American Society for Information Science and Technology 54(3), 203–215 (2003)
15. Hotho, A., Staab, S., Stumme, G.: Explaining Text Clustering Results Using Semantic Structures. In: Lavrač, N., Gamberger, D., Todorovski, L., Blockeel, H. (eds.) PKDD 2003. LNCS (LNAI), vol. 2838, pp. 217–228. Springer, Heidelberg (2003)
16. Krovetz, R., Croft, W.B.: Lexical Ambiguity and Information Retrieval (1992)
17. Manber, U.: Finding similar files in a large file system. In: Proceedings of the USENIX Winter 1994 Technical Conference on USENIX, WTEC 1994 (1994)
18. Miller, G.A.: Wordnet: a lexical database for english Commun., vol. 38. ACM (1995)
19. Sinha, R., Mihalcea, R.: Unsupervised graph-based word sense disambiguation using measures of word semantic similarity. In: ICSC, pp. 363–369 (2007)

Validation of Clustering Techniques for User Group Modeling

Danuta Zakrzewska

Abstract. Users can be grouped by clustering taking into account multiple features simultaneously. The efficiency of the method depends on the quality of obtained user groups. In the paper, the validity index, which indicates cluster quality taking into account accuracy of models, built on clusters, is considered. We assume that cluster elements are expressed by vectors of nominal values and that cluster representation is in the form of likelihood matrix. As group models frequent patterns will be considered. The performance of the proposed index will be investigated on the basis of the experiments carried out for groups of e-learning system users described by their learning styles. Comparisons to relative criteria of cluster validity will be presented.

1 Introduction

Finding groups of users of similar preferences can help in tailoring web based information system to their needs. Users may be grouped supervisively or by using clustering technique taking into account multiple features simultaneously. The efficiency of the last approach depends on the quality of obtained user groups. From the point of view of adaptation abilities of the system, cluster validity depends not only on its compactness, but also on the potential for building effective group labels, which describe traits of their members. We will say that cluster forms a "good" group of users, if it is possible to find the precise pattern of their features.

Despite of commonly used techniques, which measure compactness and separability of clusters, the validity index, which takes into account accuracy of models built on clusters will be considered. It is based on the idea of cluster representation described

Danuta Zakrzewska

Institute of Information Technology, Lodz University of Technology Wolczanska 215,
90-924 Lodz, Poland

e-mail: dzakrz@ics.p.lodz.pl

in [11]. We assume that cluster labels are not known previously and that cluster elements are expressed by vectors of nominal values, which can depict different user features. As group models frequent patterns will be pondered. The performance of the examined index will be investigated on the basis of the experiments carried out for groups of students described by their learning styles. We will consider Felder & Silverman [2] model, for which there exist explicit mappings between nominal and numerical representations of learning style dimensions. Therefore we can compare the performance of the examined index to relative validity criteria.

The paper is organized as follows. The related work concerning cluster validation is depicted in the next section together with the short description of relative criteria. In the following section the considered validity index is presented. In Section 4 the case study of student groups is depicted. In Section 5 some experimental results are described and discussed. Finally, concluding remarks and future research are outlined.

2 Cluster Validation

There exist several types of cluster validation techniques [6]. As the main criteria for cluster evaluation Jain & Koronios [5] mentioned the choice of an optimal clustering schema, as well as compactness and cohesion. The quality criteria can be divided, depending on the applied mathematical tools, into external, internal and relative. The first two approaches are based on statistical testing and are computationally complex, while the third one aims at comparing of validity index values for different clustering schemes [3]. The effectiveness of some validity indices, as well as of different clustering algorithms, for both artificial and real datasets was examined in [7]. All the validity criteria, for different clustering algorithms, have been broadly investigated in [9].

Relative criteria, which are not so computationally expensive as internal and external ones, enables the inclusion of the proper index as a part of the clustering process, and this way, help in choosing the cluster structure of the best possible quality. Their performance will be compared with the effectiveness of the considered index.

Xu & Wunsch [9] indicated relative criteria to be good tools, not only for comparing different clustering algorithms and their parameters, but also for determining the optimal number of clusters. We will consider two of the examined there indices: Dunn and SD, the ones which have chosen the majority of the optimal clustering schemes, in the case of student data (see [12]).

Let us consider a dataset O of n -dimensional objects, which are assigned to K clusters $\{SCM_j, j = 1, \dots, K\}$. Let v_j denotes the centroid of j -th cluster for $j = 1, 2, \dots, K$. Dunn's index (VD) indicates compact and well separated clusters. It was introduced by Dunn [1] and is defined as:

$$VD = \min_{i=1, \dots, K-1} \max_{j=i+1, \dots, K} \frac{d(SCM_i, SCM_j)}{\max_{i=1, \dots, K} diam(SCM_i)}, \quad (1)$$

where $d(SCM_i, SCM_j)$ is the distance between clusters SCM_i and SCM_j , measured as the minimum distance between two points belonging to SCM_i and SCM_j , respectively:

$$d(SCM_i, SCM_j) = \min_{o \in SCM_i, p \in SCM_j} d(o, p), \quad (2)$$

and the diameter $diam(SCM_i)$ of the cluster is the maximum distance between two of its members:

$$diam(SCM_i) = \max_{o, p \in SCM_i} d(o, p). \quad (3)$$

A high value of VD indicates the existence of compact and well separated clusters.

The SD validity index was introduced by Halkidi, Vazirgiannis & Batistakis [3]. It takes into account two factors: average scattering of clusters S_A and total separation between clusters S_t and is defined as:

$$SD = \alpha S_A + S_t. \quad (4)$$

The first element represents intra cluster variance measures, while the second one inter cluster distance measures and α is a weight coefficient. The average scattering of clusters S_A is constructed as follows:

$$S_A = \frac{1}{K} \sum_{i=1}^K \frac{\|\sigma(SCM_i)\|}{\|\sigma(O)\|}, \quad (5)$$

where $\sigma(O)$ and $\sigma(SCM_i)$ are vectors of variances for the dataset O and the cluster $SCM_i, i = 1, 2, \dots, K$, respectively. Let $|O|$ denotes the number of objects in the set O , then

$$\sigma(O) = [\sigma_1, \sigma_2, \dots, \sigma_n], \quad (6)$$

where

$$\sigma_i = \frac{1}{|O|} \sum_{o \in O} (o_i - SR_{O_i})^2, \quad (7)$$

$SR_O = [SR_{O_1}, SR_{O_2}, \dots, SR_{O_n}]$ is the centroid of $O, i = 1, 2, \dots, n$. Vectors of variances for clusters are calculated similarly, by taking into account cluster objects, and cluster centroids.

The total separation S_t between clusters depends on the number of clusters and is defined by using distances between cluster centroids. It is constructed as follows:

$$S_t = \frac{d_{max}}{d_{min}} \sum_{i=1}^K \left(\sum_{j=1}^K \|SR_i - SR_j\| \right)^{-1}, \quad (8)$$

where

$$d_{max} = \max \{ \|SR_i - SR_j\|, 1 \leq i, j \leq K \}, \quad (9)$$

$$d_{min} = \min \{ \|SR_i - SR_j\|, 1 \leq i, j \leq K \}, \quad (10)$$

SR_i and SR_j are centroids of clusters i and j , $1 \leq i, j \leq K$. A small value of average scattering of clusters indicate compact clusters, while total separation between clusters increases with increasing number of clusters. As both of the values are of the different range a weighting factor α , equal to total separation of clusters for the maximum number of input clusters, was proposed [3]. The number of clusters, of minimal SD index can be considered as optimal [3].

3 S_GS Validity Index

Validation of the clustering schema should depend on the aim of grouping. In the case of user groups building in order to personalise the system, big inner similarity of members of the same group is required. What is more, possibility to single out certain characteristics as being dominant in the group is very important. Such feature will enable building the explicit model of group members of the minimal error. In the case of user traits of nominal type cluster quality can be examined by using its representation (see Def. 1).

Let us assume that each user ST is described by a vector of N attributes of nominal type:

$$ST = (st_1, st_2, \dots, st_N) , \quad (11)$$

where st_i may take on k_i nominal values, $i = 1, \dots, N$. Then group representation can be defined as follows [11]:

Definition 1. Let GS be a group of objects described by vectors of N components of nominal type, each of which of M different values at most. As the group representation GSR we will consider the set of column vectors $gsr_i, i = 1, \dots, N$ of k_i components, representing attribute values of GS objects where $M = \max \{k_i : i = 1, \dots, N\}$. Each component of a vector $gsr_i, i = 1, \dots, N$ is calculated as likelihood $P_{ij}, i = 1, \dots, N; j = 1, \dots, k_i$ that objects from GS are characterized by the certain attribute value and is called the support for the respective attribute value in GS .

The group representation helps in looking at the group features comprehensively. The values of the biggest support indicate dominant attribute values for each group that can be used to build cluster models. The bigger support value means the better cluster quality. In the ideal case the group representation is constituted by vectors, which one component is equal to 1 while the others are equal to 0. Then cluster model can be determined explicitly. As the group modeling concerns all user traits, all the components of representation vectors should be taken into account simultaneously during cluster quality evaluation. Then cluster validity index based on the group representation can be expressed by a product of maximum values of group representation column vectors and can be defined as follows:

Definition 2. Let us consider the set of clusters $\{SCM_j, j = 1, \dots, K\}$ built for the objects of the dataset O . Let for every $j = 1, \dots, K$; $\{gsr_{ji}, i = 1, \dots, N\}$ be the set of column vectors of k_i components each, being representation of clusters SCM_j ,

$j = 1, \dots, K$ according to Def. 1. Then the validity index of each cluster SC_GS_j can be defined as follows:

$$SC_GS_j = \prod_{i=1}^N \max \{gsr_{j_i}[l], l = 1, \dots, k_i\}, \quad (12)$$

and respectively validity index of the clustering schema:

$$S_GS = \sum_{j=1}^K P(SCM_j) * SC_GS_j, \quad (13)$$

where $P(SCM_j)$ is computed as the ratio of the size of SCM_j and the size of the entire dataset O for $j = 1, \dots, K$.

S_GS index takes into account sizes of different groups. Thus, qualities of groups of large numbers of users have bigger influence on final index values and impact of small clusters of good qualities on a validity index value is smaller. Maximal S_GS indicates the optimal clustering schema. The degree of group separation does not have any influence on S_GS value.

Validation index S_GS can be also used in the case of clusters of users described by attributes of numerical types if there exist explicit assignments of numerical values to nominal ones. In that case effectiveness of S_GS approach can be compared to other validity criteria.

4 Student Case Study

Let us consider Felder & Silverman [2] model, which is based on *Index of Learning Style* (ILS) questionnaire, developed by Felder & Solomon [4]. The ILS is a self-scoring questionnaire for assessing preferences on 4 dimensions of the Felder & Silverman model, from among excluding pairs: *active* vs. *reflective*, *sensing* vs. *intuitive*, *visual* vs. *verbal*, *sequential* vs. *global*. The index has the form of the odd integer from the interval [-11,11], assigned for one of the dimensions from the pairs mentioned above. Each student can be modeled by a vector SL of 4 integer attributes:

$$SL = (sl_1, sl_2, sl_3, sl_4) = (l_{ar}, l_{si}, l_{vv}, l_{sg}) , \quad (14)$$

where l_{ar} means scoring for *active* (if it has negative value) or *reflective* (if it is positive) learning style, and respectively l_{si}, l_{vv}, l_{sg} are points for all the other dimensions, with negative values in cases of *sensing*, *visual* or *sequential* learning styles, and positive values in cases of *intuitive*, *verbal* or *global* learning styles.

Score from the interval [-3,3] means that the student is fairly well balanced on the two dimensions of that scale. Values -5,-7 or 5,7 mean that student learns more easily in a teaching environment which favors the considered dimension; values -9,-11 or 9,11 mean that learner has a very strong preference for one dimension of the scale and may have real difficulty learning in an environment which does not support that

preference [4]. The above principles of Felder & Silverman model impose rules for changing SL into a vector ST of nominal values. As scores from -11 to -5 (5 to 11) indicate the same favorite dimension, number of nominal values can be limited to 3. Explicit mapping between nominal values and the numerical ones allows not only to use S_GS index for student cluster validation but also to compare its performance with the ones of relative validation criteria.

5 Experiment Results and Discussion

The aim of the experiments was to evaluate the proposed validity index and to compare its performance with the results obtained by relative validity criteria. We checked modeling abilities of the clusters of the biggest SC_GS values.

For the modeling purposes dominant features of cluster members were considered. They were regarded as frequent items and used for building association rules by using Apriori-type algorithm implemented in Weka software [8]. The algorithm iteratively reduces the minimum support until it finds the required number of rules with the given minimum confidence. (Confidence of the rule $A \Rightarrow B$ is the likelihood that the itemset containing A holds also B). In all the rules found out by the algorithm operator \Rightarrow was changed to \wedge , as they were valid for the both sides. Models for the clusters of the biggest SC_GS values were investigated taking into account the accuracy measured by their support value.

The tests were done for real data of 194 Computer Science students, who filled ILS questionnaire. To obtain the most impartial results, data were gathered from students of different levels and years of studies, including part-time and evening courses. We consider clusters of disparate structures and sizes, built by application of clustering algorithms: EM, K-means, Farthest First Traversal (FFT) [8] and two-phase hierarchical [10]. Students were grouped into 3,4,5,6 and 7 clusters. In the case of K-means algorithm 2 different distance functions were used: Euclidean and Manhattan. That way 25 different clustering schemes with the total number of 125 clusters were investigated.

Table 1 presents clusters of the biggest SC_GS values. First columns concern clustering schema (technique & number of clusters), and its S_GS index. Then quantity of group members, SC_GS and models on clusters are presented. The last column contains model errors, which are calculated as the ratio of number of students for whom the model is not valid to the number of all of the students in the cluster. It can be easily noticed that presented models concern majority of the cluster members.

Table 2 presents clusters of the smallest SC_GS values and is of the same structure as the previous one. One can easily notice that model errors are much bigger than in the previous case. Many of them get close to 0.5.

S_GS index values were compared with Dunn and SD indices for all the considered clustering schemes. Table 3 presents three best schemes indicated by each of the indices.

Table 1 Clusters of the biggest SC_GS value

Method	Schema	S_GS	Quantity	SC_GS	Model	Error
Two-phase	7	0.2904	3	0.67	<i>visual</i> \wedge <i>active</i> <i>sensing</i> \wedge <i>active</i> <i>sensing</i> \wedge <i>active</i> \wedge <i>global</i>	0 0 0.33
EM	5	0.3172	14	0.6603	<i>visual</i> \wedge <i>active</i>	0
EM	7	0.3152	8	0.5808	<i>sequential</i>	0.125
FFT	7	0.1922	4	0.5625	<i>visual</i> \wedge <i>active</i> <i>sensing</i> \wedge <i>active</i> \wedge <i>visual</i> \wedge <i>global</i>	0 0.25
Kmeans (Manh)	7	0.3794	39	0.5324	<i>visual</i> \wedge <i>active</i>	0.154
Kmeans (Eucl)	7	0.2937	39	0.5149	<i>visual</i> <i>active</i>	0.077 0.2051

Table 2 Clusters of the smallest SC_GS value

Method	Schema	S_GS	Quantity	SC_GS	Model	Error
FFT	4	0.1479	38	0.0993	<i>active</i>	0.3947
FFT	3	0.1356	40	0.1017	<i>visual</i> <i>active</i>	0.375 0.425
Kmeans (Eucl)	7	0.2937	29	0.1038	<i>intuitive</i> <i>global</i> <i>visual</i>	0.3448 0.4483 0.4828
Kmeans (Eucl)	4	0.2187	37	0.1097	<i>visual</i> <i>global</i> <i>intuitive</i>	0.3784 0.4865 0.4865
EM	6	0.2518	38	0.1112	no rules	
FFT	3	0.1356	12	0.1127	<i>global</i>	0.4167
FFT	5	0.1791	28	0.1127	<i>visual</i>	0.1429

Table 3 Indices and their best schemes

Best	Method	Schema	S_GS	Dunn	SD
S_GS	Kmeans (Manh)	7	0.3794	0.1367	0.9211
	Kmeans (Manh)	6	0.3223	0.1280	0.8486
	EM	5	0.3172	0.2236	0.8303
Dunn	FFT	7	0.1896	0.5022	0.6982
	Two-phase	7	0.2904	0.4533	0.9947
	Kmeans (Eucl)	7	0.2937	0.3839	0.7087
SD	FFT	7	0.1896	0.5022	0.6982
	Kmeans (Eucl)	7	0.2937	0.3839	0.7087
	Kmeans (Eucl)	6	0.2838	0.1482	0.7335

Table 4 Clusters and their models for FFT schema

No	Quantity	SC_GS	Model	Error
1	16	0.1276	<i>sensing</i>	0.5
2	23	0.2867	<i>active</i>	0.3478
3	12	0.1256	<i>sequential</i>	0.5
4	25	0.1370	<i>visual</i>	0.16
			<i>global</i>	0.44
			<i>intuitive</i>	0.48
5	5	0.384	<i>global</i>	0
			<i>intuitive</i> \wedge <i>global</i>	0.2
6	109	0.1746	<i>visual</i>	0.2294
			<i>sensing</i>	0.4954
7	4	0.5625	<i>visual</i> \wedge <i>active</i>	0
			<i>sensing</i> \wedge <i>active</i>	0.25
			\wedge <i>visual</i> \wedge <i>global</i>	

Dunn and SD indices were consistent in two of the presented schemes. Both of them indicated clustering schema of 7 clusters built by FFT method as the best one. One cluster from the schema can be found in Table 1, with the best models, but it contains only 4 elements. Models for all the clusters of this schema are presented in Table 4.

Table 5 Clusters and their models for K-means schema

No	Quantity	SC_GS	Model	Error
1	11	0.2627	<i>global</i>	0.09
			<i>intuitive</i>	0.18
			<i>intuitive</i> \wedge <i>global</i>	0.27
2	18	0.2382	<i>active</i>	0.11
3	31	0.4877	<i>balanced</i>	
4	17	0.2617	<i>sequential</i>	0.29
5	48	0.3311	<i>visual</i>	0.08
			<i>sensing</i>	0.15
			<i>visual</i> \wedge <i>sensing</i>	0.23
6	30	0.3399	<i>visual</i>	0.03
7	39	0.5324	<i>visual</i> \wedge <i>active</i>	0.154

S_GS index indicated clustering schema of 7 clusters built by K-means algorithm with Manhattan distance function as the best (see Table 3). Again one cluster from the schema can be also found in Table 1 of the best models. All the models of this schema are presented in Table 5.

Comparison of Tables 4 and 5 shows that the errors of models, in the cluster schema indicated as optimal by S_GS index, are much lower than the ones from FFT schema chosen by Dunn and SD indices. However students assigned to the cluster number 3 of K-means schema were balanced in all the dimensions and no rules were distinguished. It is also worth mentioning, that in this schema there are three clusters with two attributes that occur together. We obtain groups of more than 10 objects, that contain *intuitive* \wedge *global*, *visual* \wedge *sensing* as well as *visual* \wedge *active* students. In FFT schema indicated by Dunn and SD indices, such situations take place only for clusters of small sizes.

6 Concluding Remarks

In the paper the problem of validity of user grouping was considered. We assumed that users can be described by nominal attributes and that they are grouped by unsupervised classification. As clusters of good quality we considered the ones, for which user models cover the majority of cluster members. It means the small error of models built on clusters. As user model association rules were considered.

The new validity index, which takes into account modeling abilities of clusters was introduced. The presented approach is based on group representation in a probability form. The proposed index allows to indicate the best clusters as well as the

optimal clustering schema. The performance of the index was examined for users of e-learning system described by their learning styles. Application of Felder & Silvermann model, which assumes explicit mappings between nominal and numerical attributes of learning styles allows to compare the proposed index with relative validity criteria.

Experiments showed that in most of the cases S_GS index indicated clusters of good modeling abilities. Model errors for optimal schema chosen by S_GS were smaller than the ones for the schema indicated by Dunn and SD indices. What is more, in the case of the last schema some student features which occur together were not distinguished.

Future research will consist in further examination of the index performance, taking into account different kinds of users and their features, including user historical behavior.

References

1. Dunn, J.C.: Well separated clusters and optimal fuzzy partitions. *J. Cybernetics* 4, 95–104 (1977)
2. Felder, R.M., Silverman, L.K.: Learning and teaching styles in engineering education. *Eng. Educ.* 78, 674–681 (1988)
3. Halkidi, M., Vazirgiannis, M., Batistakis, Y.: Quality scheme assessment in the clustering process. In: Proc. of the 4th European Conf. on Principles of Data Mining and Knowledge Discovery, Lyon, pp. 265–276 (2000)
4. ILS Questionnaire,
<http://www.engr.ncsu.edu/learningstyles/ilsweb.html>
5. Jain, R., Koronios, A.: Innovation in the cluster validating techniques. *Fuzzy Optimization and Decision Making* 7, 233–241 (2008)
6. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Comput. Surv.* 31, 264–323 (1999)
7. Maulik, U., Bandyopadhyay, S.: Performance evaluation of some clustering algorithms and validity indices. *IEEE T. Pattern Anal.* 24, 1650–1654 (2002)
8. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann Publishers, San Francisco (2005)
9. Xu, R., Wunsch II, D.: Clustering. IEEE Press & Wiley, Piscataway, NJ (2009)
10. Zakrzewska, D.: Cluster analysis in personalized e-learning systems. In: Nguyen, N.T., Szczerbicki, E. (eds.) Intelligent Systems for Knowledge Management. SCI, vol. 252, pp. 229–250. Springer, Heidelberg (2009)
11. Zakrzewska, D.: Building group recommendations in e-learning systems. In: Jędrzejowicz, P., Nguyen, N.T., Howlett, R.J., Jain, L.C. (eds.) KES-AMSTA 2010, Part I. LNCS, vol. 6070, pp. 391–400. Springer, Heidelberg (2010)
12. Zakrzewska, D.: Validation of clustering techniques for student grouping in intelligent e-learning systems. In: Jozefczyk, J., Orski, D. (eds.) Knowledge-Based Intelligent System Advancements: Systemic and Cybernetic Approaches, pp. 232–251. IGI Global (2011)

Parking Lot Occupancy Detection Using Computational Fluid Dynamics

Tomas Fabian

Abstract. In this paper, we present an algorithm for estimating the occupancy of individual parking spaces. Our method is based on a computer analysis of images obtained by a camera system monitoring the activities on a parking lot. The main idea is to use the a priori available information about the parking lot geometry and the general shape of common cars to obtain a reliable status of a parking space. We strive to avoid the training phase as much as possible to reduce the time required to bring the system into a fully operational state. To achieve this goal, we focus on a probabilistic car model and a physically based feature extraction using computational fluid dynamics. Despite the fact that the very first system of a similar type has appeared more than forty years earlier, this area is still an active research topic and a completely satisfactory solution has not been found yet.

1 Introduction

Visual surveillance systems include a wide range of the computer vision related areas. Some of the most significant areas are motion detection, moving object classification, tracking, activity understanding and semantic description. Typical applications include traffic surveillance, security, classification of activities and behaviours, the pre-crash safety system of a vehicle, and various commercial applications like license plate recognition, toll road or speeding detection systems. In addition to these common usages, there are also more specific applications, e.g. the area of the parking guidance systems. In the past decade, quite a lot of works concerned in vacant parking space detection appeared. There exist four main categories of parking

Tomas Fabian
VSB-Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science
Department of Computer Science
17. listopadu 15/2172, 70833 Ostrava-Poruba, Czech Republic
e-mail: tomas.fabian@vsb.cz

guidance systems using different technologies including the counter-based, wired-sensor-based, wireless-sensor-based and vision-based approaches [1].

The previously published methods can be divided into distinct classes that differ in the principle used to determine the parking space state. Yamada and Mizuno [12] analyzed the distribution of segments in parking cell and reported a 98.7% successful detection rate. Lin et al. [8] use an adaptive background model for each parking space. They suppose that the distribution of parking spaces is known in advance. Wu and Zhang [11] introduced a new approach in the sense of shadow and occlusion handling and this work was further improved by Wu et al. [10]. The classification is done by the general binary SVM classifier adapted by one-against-one strategy that takes all possible two-class combinations. Huang et al. [4, 5] proposed a robust parking space detection based on the three-layer Bayesian hierarchical framework. Ichihashi et al. [6, 7] introduced a parking space state classifier based on fuzzy c-means clustering and hyper parameter tuning by particle swarm optimization.

The main goal of our approach is to minimize the tremendous learning phase as much as possible. The necessity of a learning data set preparation can negatively influence the applicability in general scenarios which can consequently preclude the ease of practical implementation. Therefore, we use the a priori knowledge of the parking lot geometry and with the aim of an occlusion handling model (see Fig. 1) we obtain the normalized images of individual parking spaces (see Fig. 4) providing a set of discriminative features (see Fig. 2a) indicating the presence of a vehicle. These features are further transferred with a selective flow (see Fig. 2b) to filter out the dubious features and the final parking space state is inferred from the total weight of all features which are passing through the detection zones.

The rest of the paper is organized as follows. Section 2 presents our occlusion handling method. In Section 3, we discuss discriminative features extraction and transfer. Section 4 deals with CFD related subject of advection initialization. Weight assignment procedure is explained in Section 5. Section 6 summarizes the performance of the devised algorithm. Section 7 bears final thoughts and conclusion.

2 Occlusion Handling

In order to cope with inter-vehicle occlusions, we propose a probabilistic 3D model of a vehicle. This model represents all feasible positions of a vehicle inside the parking space. In the most simplistic way, the model can be represented by a cuboid positioned at the parking lot surface. The model is fully defined by its width, length, height, position of center and yaw. These parameters are treated as independent normally distributed random variables. As a result, we obtain a 3D scalar field of $128 \times 128 \times 256$ values representing the probability that the particular region inside the volume over a single parking space belongs to a vehicle. This can be expressed, in terms of conditional probabilities, as the likelihood $P(\mathbf{x}|\text{vehicle})$, where \mathbf{x} represents some discrete volume (voxel) inside this scalar field. To put this model in the relation with the camera, we can cast a ray through the continuous scalar field

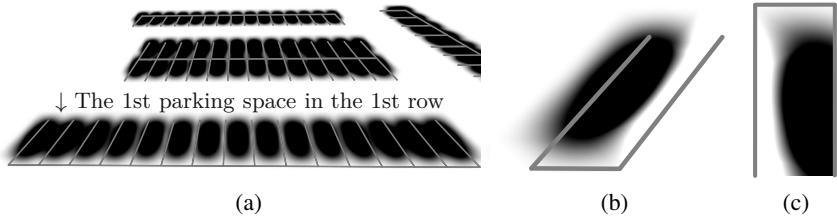


Fig. 1 (a) The scalar field h representing the parking lot surface occlusions. *Pure black level* refers to completely occluded parts of the parking lot ground. (b)-(c) The scalar field c_1 represents the confidence that the pixels may belong to a vehicle parked exclusively on the first parking space. *Black level* refers to 100% confidence

$\rho : \mathbb{R}^3 \rightarrow \langle 0, 1 \rangle$ which is obtained as a trilinear interpolation of the three-dimensional discrete field of likelihoods $P(\mathbf{x}|\text{vehicle})$. With the aim of basic calculus, we can formulate the expression for the scalar field of occlusions h in terms of the line integral along a piecewise smooth curve L (line of view made up of the set of straight segments intersecting affected voxels V) as follows

$$\begin{aligned} h(x, y) &= \bigcup_{\mathbf{x} \in V} P(\mathbf{x}|\text{vehicle}) = \int_L \rho(s) \, ds = \int_a^b \rho(\mathbf{r}(t)) \|\mathbf{r}'(t)\| \, dt \\ &= \int_a^b \rho(\mathbf{r}(t)) \|\hat{\mathbf{d}}\| \, dt = \int_a^b \rho(\mathbf{r}(t)) \, dt = \dots = \sum_{i \in I} \int_{t_i}^{t_{i+1}} \sum_{j=0}^3 a_j t^j \, dt, \quad (1) \end{aligned}$$

where the ray $\mathbf{r}(t) = O + \hat{\mathbf{d}}t$ is a bijective parameterization of the line segment originating at the point $\mathbf{r}(a)$ coincident with camera's origin O and the end point $\mathbf{r}(b)$ is the intersection with the parking lot plane. In addition, the integral over the interval $\langle a, b \rangle$ is decomposed into the sum of integrals over the set I of intervals $\langle t_i, t_{i+1} \rangle$ representing the parametric coordinates of intersections of the ray \mathbf{r} with the set of affected voxels V . The analytical derivation of parameters a_0, a_1, a_2 and a_3 is relatively straightforward but tedious involving more than one hundred summands in the final expression. We left them out of this calculation to avoid unnecessary clutter. Equation (1) can be also solved numerically by probabilistic algorithms such as Monte-Carlo method. Returned scalar value of the function h represents the degree of our belief that the certain position (x, y) in the image of parking lot surface can be occluded exclusively by a well-parked car (see Fig. 1a). The resulting confidence field for the i -th parking space equals to $c_i(x, y) = 2h_i(x, y) - h(x, y)$, where h_i is occlusion map where only i -th parking place is occupied and h is occlusion map generated for the fully engaged parking lot. Figure 1b presents the unprojected confidence field for the first parking space and in Fig. 1c is the projected version of the same field.

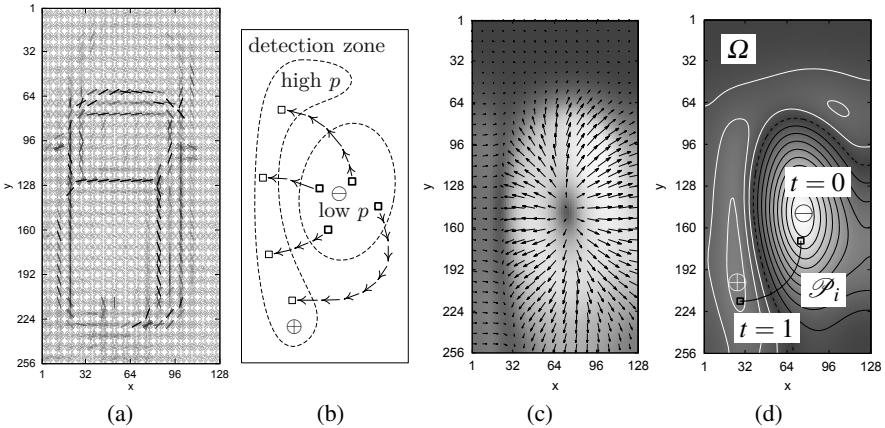


Fig. 2 (a) The plot shows the result of HOG features extraction. (b) The scheme depicts expected behaviour of features advection. Features are transported from low pressure areas into high pressure areas (detection zones). (c) The plot shows an example of the initial force field \mathbf{f} after 10000 iterations. (d) The high-pressure field marks the detection areas along the car boundaries (white *iso* lines)

3 Image Features Extraction and Transfer

At this point, we have obtained the rectangular image of every parking place, the related confidence field and we would like to extract the relevant features that would allow us to discriminate between the two possible states of parking spaces. Local object appearance and shape can often be characterized by the distribution of local intensity gradients or edge directions. Dalal and Triggs [2] showed in their experiments that the Histogram of Oriented Gradient (HOG) is one of the most successful edge and gradient-based descriptor and significantly outperforms existing feature sets for human or car detection.

Having the information about the distribution of local gradients in the normalized image of parking space represented by the HOG features, we try to lay a foundation for a discriminative model for two-class labeling problem of individual parking places. If we look at the Fig. 2a and when we compare it with the image of the empty parking place, there is obvious difference in the distribution of prevailing edges. Simply put, the pertinence to the given class may be devised from the total amount of cells, which can be regarded as the parts of parked car edges. Here arises the problem of how to determine which cell belongs to an car edge. We assume that the potential car edges can appear in the area restricted by the confidence field of every parking place. Moreover, based on the vector field \mathbf{n} obtained by projection of three-dimensional vector field of the car model iso-surface normals (e.g. for $h = 1$) onto the parking space normalized image plane, we can roughly estimate the direction of such edges (i.e. the expected edge will be perpendicular to the local normal vector). In real world, the distribution of edges caused by the presence of a car will

be somewhat more irregular than suggested by that geometrical model, yielding the need to handle such variances.

Physically-inspired from the classical fluid dynamics, we may think of every cell from the Fig. 2a as an idealized flowing fan-like mass object. Cells more conformal with the edge model will experience stronger drag force resulting in a higher velocity of these cells. As a result, these cells will be easily advected by the flow field from the origin position into the detection zone. On the opposite side, uniformly distributed bins will resist the flow. This will introduce the desired flexibility of our discrimination model with respect to the underlying car edge model. The here described idea is visualized in the Fig. 2b. The motion of a Newtonian fluid with a constant density and temperature is governed by the Navier-Stokes equations (NSE) as follows

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{u} + \mathbf{f}, \quad (2)$$

where \mathbf{u} represents the velocity vector field, p is the pressure field, ρ is the fluid density and ν is the kinematic viscosity of the fluid. The vector field \mathbf{f} is an external force field and will be discussed in the following Section 4. In the case of incompressible fluids, the conservation of mass is then stated as the continuity equation $\nabla \cdot \mathbf{u} = 0$ meaning that the divergence of vector field \mathbf{u} is zero. For the sake of brevity, the implementation details of solving the NSE are left uncovered. We adopt the approach thoroughly described in [9].

The NSE were successfully applied in many fields including image analysis and in our case, we interpret the resulting pressure and velocity fields as follows. The low pressure areas correspond to the sources of strong gradients caused by eventual car edges located at the positions predicted by the field of projected normals. The high pressure regions will represent the traps for moving particles. If the particle arrive in the detection area and has a strong dominant bin in the HOG, then we can suppose that the origin of the particle is placed somewhere close to a strong edge of a parked vehicle. The trajectory \mathcal{P} in the conjunction with the actual bins configuration in particle's HOG should influence the speed of the moving particle. The new position \mathbf{r} of the cell with the total mass m in the particular time step Δt is given by the formula $\mathbf{r}^{t+1} = \mathbf{F}^t \Delta t^2 / m + 2\mathbf{r}^t - \mathbf{r}^{t-1}$. The steady-state drag force \mathbf{F} on the cell due to the fluid flow is derived from the standard quadratic drag equation for an object moving through a fluid at relatively large velocity and equals to

$$\mathbf{F}^t = \frac{1}{2} \rho \|\mathbf{u}\| \sum_{i=0}^{b-1} [1 - |\text{rad2grad}(\text{bin2rad}(i) + \sum_{j=0}^t \varphi^j) \cdot \hat{\mathbf{u}}|] C(\text{hog}(i)), \quad (3)$$

where the function `rad2grad` translates the angle in radians to unit direction vector and `bin2rad` converts i -th bin to radians. The constant b represents number of bins per orientation histogram. The hat over the \mathbf{u} means that it is a unit vector and has magnitude equal to 1. The drag coefficient C is associated with the number of votes $v_i := \text{hog}(i)$ in the i -th histogram channel through simple polynomial function

$C(v_i) = \alpha v_i^\beta$. For the rest of our experiments, the parameters were set as follows: $\alpha = 3$ and $\beta = 5$. The angle φ compensates the yaw of the cell as it moves across the simulation domain. The yaw is computed iteratively as follows

$$\varphi^{t+1} = \text{sgn}(\hat{\mathcal{P}}'(t) \times \hat{\mathcal{P}}'(t+1)) \arccos(\hat{\mathcal{P}}'(t) \cdot \hat{\mathcal{P}}'(t+1)), \quad (4)$$

starting with $\varphi^0 = \varphi^1 = 0$. The unit tangent vectors $\hat{\mathcal{P}}'$ are replaced by the backward difference approximation up to the fourth-order of accuracy.

4 Force Field Generation

We expect that the external velocity field will start transferring the features from the regions of their abundance into the detection areas (see Fig. 2b). We can start with the gradient of the iso-surface of the h function which is subsequently projected on every parking place yielding a 2D vector field of normals \mathbf{n} . The original normal vector field \mathbf{n} is very close to fulfil the stated requirements on the field \mathbf{f} which will initiate the motion of cells during the CFD steps. In order to assure that the force field fulfil the stated requirements, we define the force field to be the vector field $\mathbf{f}(x,y) = [u(x,y), v(x,y)]$ that minimizes the global energy functional

$$\begin{aligned} \mathcal{E} = \iint_{\Omega} & \alpha (\|\nabla u\|^2 + \|\nabla v\|^2) + \beta \|\mathbf{n}\|^2 \|\mathbf{f} - \mathbf{n}\|^2 \\ & + \|\nabla c\|^2 \|\mathbf{f} - \nabla c\|^2 dx dy, \end{aligned} \quad (5)$$

where the first term in the functional follows a standard principle, that of making the result smooth when there is no data. The second term is the data attachment term, whose minimization tends to make the force field to be similar with the normal field especially in the areas where the normal field is large. The third confidence field driven term enforces the presence of an in-flow from border areas and also partially helps to increase the pressure in the detection areas. After applying the standard methods of the variation calculus we obtain two Euler-Lagrange equations

$$\begin{aligned} \beta \|\mathbf{n}\|^2 (u - m) + \|\nabla c\|^2 (u - p_x) - \alpha \Delta u &= 0, \\ \beta \|\mathbf{n}\|^2 (v - n) + \|\nabla c\|^2 (v - p_y) - \alpha \Delta v &= 0, \end{aligned} \quad (6)$$

where Δ is the Laplace operator. Both Eqs. (6) can be solved iteratively by treating u and v as functions of time t according the time-marching scheme. These equations are decoupled, and therefore can be solved as separate scalar partial differential equations in u and v , provided that the partial derivatives with respect to time t on the left side of the Eqs. (7) are approximated by the first-order accurate forward difference formulas yielding

$$\begin{aligned}
u^{t+1}(x,y) &= u^t(x,y) + \Delta t \left(\beta \|\mathbf{n}(x,y)\|^2 (u^t(x,y) \right. \\
&\quad \left. - m(x,y)) + \|\nabla c(x,y)\|^2 (u^t(x,y) - c_x(x,y)) - \alpha \Delta u^t(x,y) \right), \\
v^{t+1}(x,y) &= v^t(x,y) + \Delta t \left(\beta \|\mathbf{n}(x,y)\|^2 (v^t(x,y) \right. \\
&\quad \left. - n(x,y)) + \|\nabla c(x,y)\|^2 (v^t(x,y) - c_y(x,y)) - \alpha \Delta v^t(x,y) \right).
\end{aligned} \tag{7}$$

The iteration begins by setting $u^0(x,y) = m(x,y)$ and $v^0(x,y) = n(x,y)$. To ensure the convergence of the above described iterative process, we restrict the time step Δt with the Courant-Friedrichs-Lowy (CFL) condition. An example of the resulting force field is shown in the Fig. 2c.

5 Weight Assignment Procedure

In this section we will describe how to assign certain weight w_i to the individual cell (i.e. the rate of belonging to the car edge). As stated above, we track the position \mathbf{r} of every cell as it moves across the simulation domain Ω represented by the normalized image of a parking space. In accordance with our cells advection model, the most relevant cells travel across high-pressure areas and should gain the most votes (or weight). This can be expressed by the following path integral

$$w_i = \int_{\mathcal{P}_i} \kappa(p(s)) ds = \int_0^1 \kappa(p(\mathbf{r}(t))) \|\mathbf{r}_t(t)\| dt, \tag{8}$$

where \mathcal{P}_i is the trajectory taken by the i -th cell due to the influence of the flow field \mathbf{u} (see Fig. 2d). The factor $\|\mathbf{r}_t(t)\|$ represents the speed of traversal of the trajectory

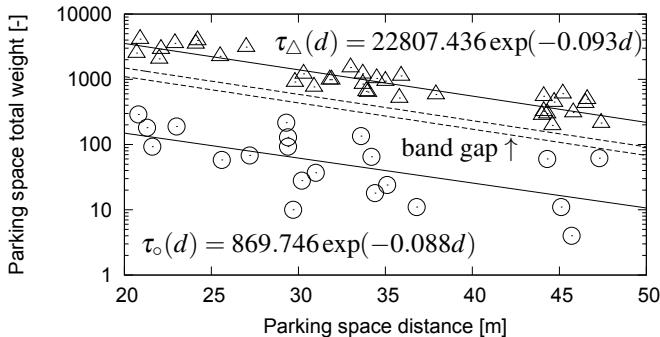


Fig. 3 The plot with a logarithmic scale specified for the vertical axis shows the dependence of the parking space total weight $\sum_{\Omega} w_i$ on the parking space distance d from the camera. Two *solid lines* represent exponential regression models of measured total weights of vacant (*circles*) and occupied (*triangles*) parking spaces

as the parameter t runs between two endpoints $t = 0$ and $t = 1$. The real-valued function κ converts the pressure into the weight value. We suggest to define this function as $\kappa(x) = \sqrt{\max(x, 0)}$. This definition reflects our exclusive interest in areas with positive pressure and also reduces the influence of the pressure magnitude on the resulting weight. Moreover, the influence of camera distance on the HOG must be compensated to retain the possibility to discriminate the parking space state with only a single threshold. The actual threshold value depends on the mean of two exponential regression models approximating the scatter plot of both categories (see Fig. 3).

6 Evaluation

The reliability of our method was evaluated under different lighting conditions and it is summarized in the Table 1. We observed 56 parking spaces in 4 rows with the network HD camera directly attached to the dedicated video server via Wi-Fi.

Table 1 Results of our algorithm compared against the ground truth data

Original image	Row	TP	FN	FP	TN	In	Out ¹	Acc.	Recall	Prec.	Spec.	F1
	1	600	0	18	1179	1	8	0.990	1.000	0.971	0.985	0.985
	2	494	0	1	1065	0	5	0.999	1.000	0.998	0.999	0.999
	3	609	0	3	948	2	8	0.998	1.000	0.995	0.997	0.998
	4	1230	12	14	539	7	7	0.986	0.990	0.989	0.975	0.990
	All	2933	12	36	3731	10	28	0.993	0.996	0.988	0.990	0.992
φ compensation		2937	9	37	3729			0.993	0.997	0.988	0.990	0.992
	1	915	0	65	817	18	3	0.964	1.000	0.934	0.926	0.966
	2	1204	29	1	316	15	2	0.981	0.976	0.999	0.997	0.988
	3	1268	0	0	292	12	2	1.000	1.000	1.000	1.000	1.000
	4	1515	0	18	267	15	3	0.990	1.000	0.988	0.937	0.994
	All	4902	29	84	1692	60	10	0.983	0.994	0.983	0.953	0.989
	1	666	0	20	205	12	0	0.978	1.000	0.971	0.911	0.985
	2	610	1	2	166	9	0	0.996	0.998	0.997	0.988	0.998
	3	525	1	4	238	10	1	0.993	0.998	0.992	0.983	0.995
	4	746	0	4	147	9	1	0.996	1.000	0.995	0.974	0.997
	All	2547	2	30	756	40	2	0.990	0.999	0.988	0.962	0.994

¹ Number of incoming vehicles during the test period.

² Number of outbound vehicles during the test period.

Individual test images contain more than 10000 occupied parking spaces and more than 6000 vacant parking spaces. In addition to the table, false positive rate (FPR) and false negative rate (FNR) are no worse than 0.047 and 0.006, respectively. In comparison, other authors report the final false acceptance rate (FAR) 0.032 and false rejection rate (FRR) 0.020 [5]. Other method based on a SVM classifier achieves false detection rate (FDR) 0.048 and FRR 0.071 [3]. Furthermore, the results point out that the yaw angle φ compensation procedure is not the critical part of our algorithm. All of these experiments were done on Intel Xeon X3220 running at 2.4 GHz and on average, it took 29 seconds to process a single frame with mostly unoptimized C++ code. The ground truth data were assessed manually by a human operator.

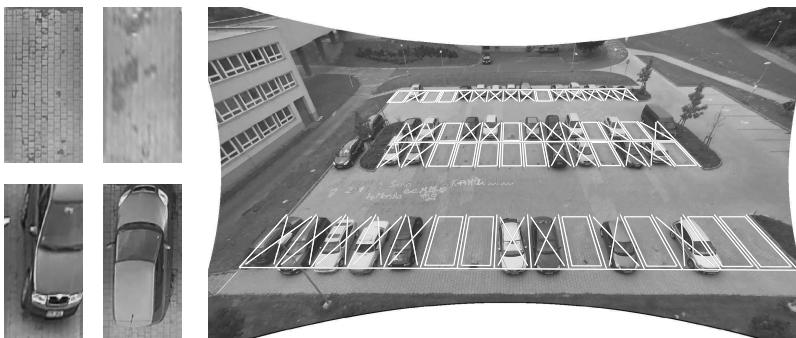


Fig. 4 The result of a single image analysis. On the left, the figure contains four normalized images of parking spaces from different parking rows.

7 Conclusion

In this paper, we presented a new algorithm for parking lot occupation detection based on the analysis of the arrangement of HOG features in the image of a single parking space. The selective flow improves the detection accuracy by filtering out the cells which seems to be unrelated with the car model and, for example, represent only the ground pattern. The experiments show that the algorithm performs well over the wide range of lighting conditions and the achieved accuracy was 96.4% in the worst case. The algorithm may be extended by enforcing contextual constraints by the Conditional Random Field (CRF). This will replace the contemporary simplified process of parking space total weight evaluation when the neighbourhood of cells is not taken in account. In our future work, we also plan to address an effective utilization of GPUs to reduce the overall latency of our parking lot surveillance system.

References

1. Bong, D., Ting, K., Lai, K.: Integrated approach in the design of car park occupancy information system (COINS). *IAENG Int. J. of Comp. Science* 35(1), 7–14 (2009)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR), San Diego, California, USA, pp. 886–893 (2005)
3. Dan, N.: Parking management system and method (2003),
http://www.patentlens.net/patentlens/patent/US_7062320/
4. Huang, C.C., Wang, S.J.: A hierarchical bayesian generation framework for vacant parking space detection. *Transactions on Circuits and Systems for Video Technology* 20(12), 1770–1785 (2010)
5. Huang, C.C., Wang, S.J., Chang, Y.J., Chen, T.: A bayesian hierarchical detection framework for parking space detection. In: Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, Nevada, USA, pp. 2097–2100 (2008)
6. Ichihashi, H., Katada, T., Fujiyoshi, M., Notsu, A., Honda, K.: Improvement in the performance of camera based vehicle detector for parking lot. In: Proc. Int. Conf. on Fuzzy Systems (FUZZ), Barcelona, Spain, pp. 1–7 (2010)
7. Ichihashi, H., Notsu, A., Honda, K., Katada, T., Fujiyoshi, M.: Vacant parking space detector for outdoor parking lot by using surveillance camera and fcm classifier. In: Proc. Int. Conf. on Fuzzy Systems (FUZZ), Jeju Island, Korea, pp. 127–134 (2009)
8. Lin, S.F., Chen, Y.Y., Liu, S.C.: A vision-based parking lot management system. In: Proc. Int. Conf. on Systems, Man and Cybernetics (SMC), Taipei, Taiwan, pp. 2897–2902 (2006)
9. Stam, J.: Stable fluids. In: Proc. 26th Annual Conf. on Computer Graphics (SIGGRAPH), Los Angeles, California, USA, pp. 121–128 (1999)
10. Wu, Q., Huang, C., Yu Wang, S., Chen Chiu, W., Chen, T.: Robust parking space detection considering inter-space correlation. In: Proc. Int. Conf. on Multimedia and Expo (ICME), Bejing, China, pp. 659–662 (2007)
11. Wu, Q., Zhang, Y.: Parking lots space detection. Tech. rep., Carnegie Mellon University, Pittsburgh, Pennsylvania, USA (2006)
12. Yamada, K., Mizuno, M.: A vehicle parking detection method using image segmentation. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* 84(10), 25–34 (2001)

Human Fall Detection Using Kinect Sensor

Michał Kepski and Bogdan Kwolek

Abstract. Falls are major causes of mortality and morbidity in the elderly. The existing CCD-camera based solutions require time for installation, camera calibration and are not generally cheap. In this paper we show how to achieve automatic fall detection using Kinect sensor. The person is segmented on the basis of the updated depth reference images. Afterwards, the distance of the person to the ground plane is calculated. The ground plane is extracted by the RANSAC algorithm. The point cloud belonging to the floor is determined using v-disparity images and the Hough transform.

1 Introduction

Falling is an everyday possible accident that all of us are exposed to. A fall can lead to severe consequences, such as fractures, and a fallen person might need assistance at getting up again. Thus, in recent years a lot of research has been dedicated into the development of fall detection methods [16] [14]. Such methods are designed to robustly detect falls and then to raise a medical alert. Medical personnel can then be dispatched to the site where the alarm was activated.

As humans become old, their bodies weaken and the risk of accidental falls raises noticeably [11]. The research results demonstrate that high percentage of injury-related hospitalizations for seniors are the results of falls [6]. Since the population of elderly is increasing dramatically in almost all countries of the world, high demand for unobtrusive and assistive technology is observed. In particular, the assistive technology can contribute toward independent living of the elderly [3] [15]. However,

Bogdan Kwolek

AGH University of Science and Technology, 30 Mickiewicza Av., 30-059 Krakow, Poland

Michał Kepski

University of Rzeszów, 16c Rejtana Av., 35-959 Rzeszów, Poland

regardless of numerous efforts undertaken to attain reliable and unobtrusive fall detection, current technology does not meet the requirements of the seniors [20]. False alarms can happen while seniors are bending over, laying down, or doing a variety of other day-to-day activities. In general, current technology leads to much higher rate of false alarms when compared with standard medical alert.

Applicable techniques for fall detection include a variety of methods. Most of the techniques are based on body-worn or built-in devices, which are intrusive as they require the user to wear a smart device. Such methods utilize accelerometers or both accelerometers and gyroscopes to discriminate the fall from activities of daily living (ADLs) [15]. However, very often on the basis of such sensors it is hard to separate real falls from fall-like activities [2]. Bending over, laying down, sitting down, or even setting down a purse can all resemble a fall depending on how it was done. In consequence, these methods trigger significant number of false alarms. What's more, the detectors that are typically attached to a belt around the hip, are uncomfortable to be worn during the sleep [5]. Furthermore, their usefulness in monitoring of critical phases like getting up from the bed is relatively poor. In addition to applications presented in scientific publications, commercial fall detection systems are available as shown in a survey [16] with 7 examples of commercially available fall detection systems and over 40 patents on fall detectors.

During the recent years, a lot of research has been done on detecting falls using a wide range of sensor types [15] [20], including pressure pads [19], single CCD camera [1] [18], multiple cameras [4], specialized omni-directional ones [13] and stereo-pair cameras [7]. Video cameras offer several advantages over other sensors including the capability of detection of various activities. The further benefit is low intrusiveness and the possibility of remote verification of fall events. However, the currently available solutions require time for installation, camera calibration and they are not generally cheap. As a rule, CCD-camera based systems require a PC computer or a notebook for image processing. The existing video-based devices for fall detection cannot work in nightlight or low light conditions. Additionally, the lack of depth information can lead to lots of false alarms. Moreover, in most of such systems the privacy is not preserved adequately.

Recently, the Kinect sensor has been successfully used in fall detection systems [12] [17] [8]. It is the world's first system that at reasonable price combines an RGB camera and a depth sensor. Unlike 2D cameras, the low-cost Kinect allows tracking the body movements in 3D. Thus, if only depth images are used it can guarantee the person's privacy. The Kinect sensor is independent of external light conditions, since it is equipped with an active light source. As the Kinect uses infrared light it is able to extract depth images in a room that is dark to our eyes.

2 Motivation and Background

Depth is very useful cue to achieve reliable person detection because humans may not have consistent color and texture but have to occupy an integrated region in space. However, in many home scenarios is not easy to detect a person using only

depth images due to occlusions, for instance, if a person stands behind a chair being in turn in the front to the Kinect. The software called NITE, which is a binary distribution from PrimeSense offers skeleton tracking with the Kinect sensor. However, this software has been developed for human computer interaction, and not to detect the person fall. Thus, in some circumstances it has difficulties in extracting and tracking the skeleton, see Fig. 1a-b, as well as segmenting the person, see Fig. 1c, where we can see two segments belonging to the person lying on the floor.

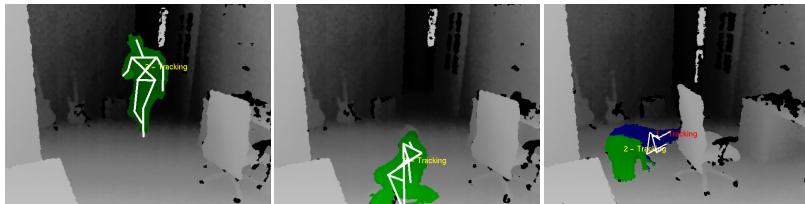


Fig. 1 NITE-based skeleton tracking during a person fall

Because of the inconveniences mentioned above as well as the lack of the distribution of the NITE for an embedded platform we elaborated algorithm for person extraction in depth images at relatively low computational cost. In order to achieve reliable fall detection we employ both Kinect and accelerometer that complement one another [8]. We implemented the system on PandaBoard ES, which is a low-power, low-cost single-board computer development platform based on Texas Instruments OMAP4 line of processors [9]. It enables development of mobile applications. Regarding low-cost computational power of the board the person was detected on the basis of the scene reference image, which was extracted in advance. Such a fast method of person segmentation can be applied in many scenarios, for instance in fall detection systems mounted on the stairs. However, in home environments such an approach can be impractical. The main reason for this is that in case of the moved furniture, like chair or even opening the door the scene reference image contains such objects, what in turns can lead to difficulties in segmentation of the person on the basis of depth connected components.

In this work we demonstrate a method for updating the depth reference image at a low computational cost. We also demonstrate how to extract the ground plane in the depth images. The ground plane is extracted automatically using the v-disparity images, Hough transform and the RANSAC algorithm.

3 Fall Detection on Embedded Platform

In this section we present the main ingredients of our embedded system for human fall detection [9]. Our fall detection system uses both data from Kinect and motion data from a wearable smart device containing accelerometer and gyroscope sensors. Data from the smart device (Sony PlayStation Move) are transmitted wirelessly

via Bluetooth to the PandaBoard on which the signal processing is done, whereas Kinect is connected via USB. The device contains one tri-axial accelerometer and a tri-axial gyroscope consisting of a dual-axis gyroscope and a Z-axis gyroscope. The fall alarm is triggered by a fuzzy inference engine based on expert knowledge, which is declared explicitly by fuzzy rules and sets. As inputs the engine takes the acceleration, the angular velocity and the distance of the person's gravity center to the altitude at which the Kinect is placed. The acceleration's vector length is calculated using data provided by the tri-axial accelerometer, whereas the angular velocity is provided by the gyroscope. The sampling rate of both sensors is equal to 60 Hz. The sensor is typically attached to trunk or lower back because such body parts represent the major component of body mass and move with most activities.

The Kinect sensor captures depth and color images simultaneously at a frame rate of about 30 fps. It consists of an infrared laser-based IR emitter, an infrared camera and a RGB camera. The IR camera and the IR projector form a stereo pair with a baseline of approximately 75 mm. Kinect depth measurement is based on structured light, making a triangulation between the dot pattern emitted and the one captured by the IR CMOS sensor. Pixels in the provided depth images indicate calibrated depth in the scene. Kinect's field of view is fifty-seven degrees horizontally and forty-three degrees vertically. The minimum range for the Kinect is about 0.6 m and the maximum range is somewhere between 4-5 m.

The depth images are acquired using OpenNI (Open Natural Interaction) library. A mean depth map is extracted in advance to delineate the foreground object at low-computational cost. It is extracted on the basis of several consecutive depth images without the subject to be monitored and then it is stored for the later use in the detection mode. In the detection mode the foreground objects are extracted through differencing the current image from such a reference depth map. Afterwards, the foreground object is determined through extracting the largest connected component in the thresholded difference map. According to the reports of the code profiler the module responsible for detection of the foreground object uses 50% of the CPU's computational power.

4 V-disparity Based Ground Plane Detection

The v-disparity images were originally proposed in [10] to achieve obstacle detection using disparity maps between two stereo images. Given a depth map extracted by the Kinect sensor, the disparity d can be calculated in the following manner:

$$d = \frac{b \cdot f}{z} \quad (1)$$

where z is the depth (in meters), b is the horizontal baseline between the cameras (in meters), f is the (common) focal length of the cameras (in pixels). For the Kinect sensor the value of b is 7.5 cm and it is the measured distance between the IR and projector lenses, whereas f is equal to 580 pixels.

Let H be a function of the disparities d such that $H(d) = I_d$. The I_d is the v-disparity image and H accumulates the pixels with the same disparity from a given line of the disparity image. Thus, in the v-disparity image each point in the line i represents the number of points with the same disparity occurring in the i -th line of the disparity image. Figure 2c illustrates the v-disparity image that corresponds to the depth image depicted on Fig. 2b. The size of the images acquired by Kinect is 640×480 .

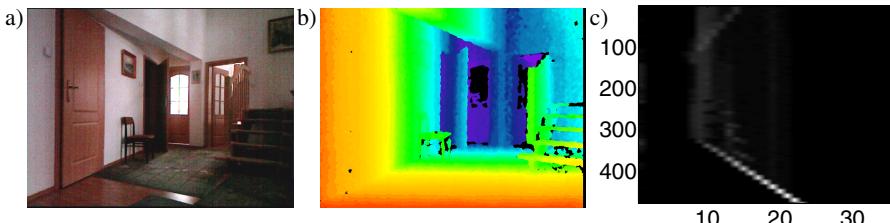


Fig. 2 V-disparity map calculated on depth images from Kinect: RGB image a), corresponding depth image b), v-disparity map c)

The line corresponding to the floor pixels was extracted using the Hough transform. Assuming that the Kinect is placed at height about 1 m from the floor, the extracted line should begin in the disparities ranging from 21 to 25 depending on the tilt angle of the sensor. On the basis of the extracted line the pixels belonging to the floor areas were determined. Due to the measurement inaccuracies we considered also pixels in some disparity extent d_t as belonging to the ground. Assuming that d_y is a disparity in the line y , which represents the pixels belonging to the ground, we take into account the disparities from the range $d \in (d_y - d_t, d_y + d_t)$ as representing the ground. Figure 3a illustrates the point cloud corresponding to images shown on Fig. 2b. On Fig. 3b we can observe the point cloud without the points belonging to the floor. Given the line extracted by the Hough transform, the points on the v-disparity image with the corresponding depth pixels were selected, and then transformed to point cloud, see Fig. 3c depicting the points cloud of the floor, which was selected using the v-disparity map.

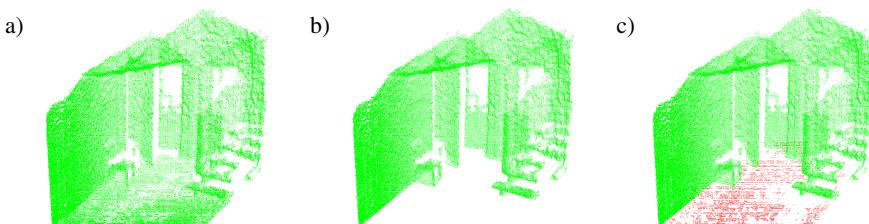


Fig. 3 Points cloud corresponding to the depth image (from Fig. 2b) a), the points cloud without the points belonging to the floor b), the points cloud from image b), with points belonging to the floor c)

After the transformation of the pixels to the 3D points cloud representing the floor, the plane described by the equation $ax + by + cx + d$ was recovered. The parameters a, b, c and d were estimated using RANSAC algorithm. The distance of the 3D centroid of the segmented person to the ground plane was determined on the basis of the following equation:

$$D = \frac{|aX_c + bY_c + cZ_c + d|}{\sqrt{a^2 + b^2 + c^2}} \quad (2)$$

where X_c, Y_c, Z_c are coordinates of the centroid.

The algorithm for ground plane extraction was tested on variety of images. Figure 4 depicts some results that were obtained on NYU Depth Dataset¹.

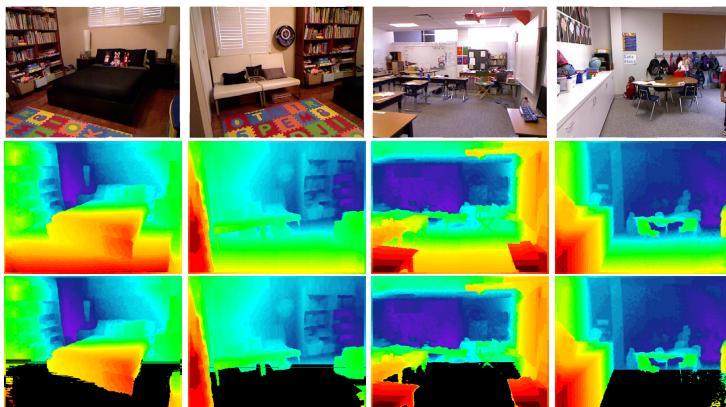


Fig. 4 Ground plane extraction. RGB images (upper row), depth images (middle row), depth images without the floor (bottom row).

5 Person Segmentation

In our previous work [9], a depth reference map-based extraction of the person has been utilized. The method has been investigated mainly due to limited computational power of the PandaBoard at which the system has been implemented. The depth reference map was extracted on the basis of several consecutive depth images without the subject to be monitored and then it was stored for the later use in the person detection mode. In the detection mode the foreground objects were extracted through differencing the current image from such a reference depth map. Experimental findings demonstrated that such a technique can be applied in many scenarios, for instance in systems monitoring fall detection on stairs. However, in case of change of the scene layout, for example due to change of furniture settings

¹ Sequences obtained from <http://cs.nyu.edu/~{}silberman/datasets/>

some non-person objects can appear in the reference images and this in turn can lead to difficulties in segmenting the person.

In order to make the system applicable in a wide range of scenarios we elaborated a fast method for updating the depth reference image. In our approach, each pixel of the depth reference image assumes the median value of several pixels from the past images. At the beginning we collect a number of images, and for each pixel we assemble a list of the pixels from the former images, which is then sorted in order to extract the median. For images of size 640×480 the computation time needed for extraction of the median is about 9 milliseconds at 2.4 GHz I7 processor running 4 threads. At the PandaBoard this operation can be completed in 0.15 sec. Given the sorted lists of pixels the depth reference image can be updated quickly by removing the oldest pixels and updating the sorted lists with the pixels from the current depth image and then extracting the median value. We found that for typical human motions, good results can be obtained using 13 depth images. For Kinect acquiring images at 25 Hz we take every tenth image.

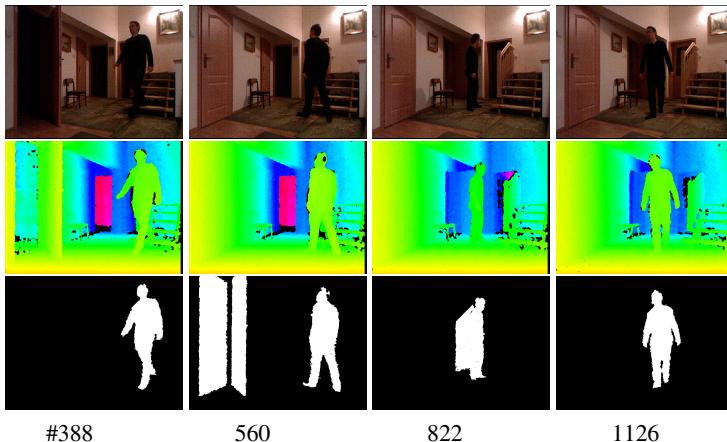


Fig. 5 Person segmentation. RGB images (upper row), the depth images (middle row) and the segmented person (bottom row).

Figure 5 illustrates some example depth reference images, which have been obtained using the discussed technique. In the image #388 we can see the opened door, which was closed to demonstrate how the algorithm updates the reference image. In frame #560 we can see that the door appears in the reference image, and then it is removed in frame #822. As we can observe, the updated reference image is free of clutter and allows us to extract the depth silhouette. In order to eliminate small objects the depth connected components were extracted. Afterwards, small artifacts were eliminated. Otherwise, the depth images can be cleaned using morphological erosion. When the person does not move the reference image is not updated. It is worth noting that the accelerometer can support the detection of periods in which the movement of the person takes place. As we utilize only one Kinect, an

occlusion happening because of the furniture (e.g. sofa or chair) can lead to difficulties in detecting the fall, when only images are employed. In such situations the decision can be made on the basis of the information provided by the accelerometer and the gyroscope.

6 Experimental Results

Four volunteers with age over 26 years attended in experiments and tests of our algorithm. A data-set of normal activities like walking, sitting down and crouching down has been composed in order to determine the threshold value, i.e. the distance of the centroid of the person to the floor below which the alarm should be triggered. Figure 6 depicts the distance D to the ground plane that has been obtained for some daily activities. As we can observe, on the basis of analysis of the centroid motion it is possible to discard some short-term actions, like sitting down, for which the centroid was temporally below the alarm threshold.

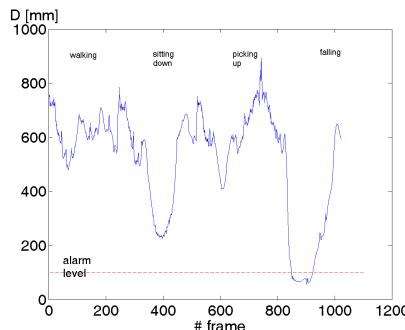


Fig. 6 Distance of the centroid to the ground plane for person performing some daily activities

Intentional falls were performed in home towards a carpet with thickness of about 2 cm. Each individual performed three types of falls, namely forward, backward and lateral at least three times. Figure 7 depicts a person who has fallen and the corresponding binary map, which was obtained through differencing the current depth image from the reference depth image and then thresholding the difference image. All intentional falls were detected correctly.

The system was implemented in C/C++ and runs at 25 fps on 2.4 GHz I7 (4 cores, Hyper-Threading) notebook powered by Linux. We are planning to implement the system on the PandaBoard.

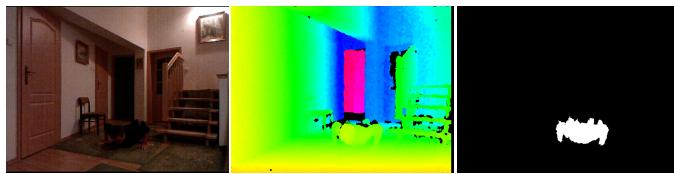


Fig. 7 Image with a fallen person (left), corresponding depth image (middle) and binary image with the extracted person (right)

7 Conclusions

In this work we demonstrated our approach to fall detection using Kinect. The detection of the fall is done on the basis segmented person in the depth images. The segmentation of the person takes place using updated depth reference images of the scene. The distance of the centroid of the segmented person to the ground plane is used to trigger the fall alarm. The ground plane is extracted automatically using the v-disparity images, Hough transform and the RANSAC algorithm.

Acknowledgements. This work has been supported by the National Science Centre (NCN) within the project N N516 483240.

References

1. Anderson, D., Keller, J., Skubic, M., Chen, X., He, Z.: Recognizing falls from silhouettes. In: Annual Int. Conf. of the Engineering in Medicine and Biology Society, pp. 6388–6391 (2006)
2. Bourke, A., O'Brien, J., Lyons, G.: Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait & Posture* 26(2), 194–199 (2007)
3. Cook, A., Hussey, S.: Assistive Technologies: Principles and Practice, 2nd edn., Mosby (2002)
4. Cucchiara, R., Prati, A., Vezzani, R.: A multi-camera vision system for fall detection and alarm generation. *Expert Systems* 24(5), 334–345 (2007)
5. Degen, T., Jaeckel, H., Rufer, M., Wyss, S.: Speedy: A fall detector in a wrist watch. In: Proc. of IEEE Int. Symp. on Wearable Computers, pp. 184–187 (2003)
6. Heinrich, S., Rapp, K., Rissmann, U., Becker, C., König, H.H.: Cost of falls in old age: a systematic review. *Osteoporosis International* 21, 891–902 (2010)
7. Jansen, B., Deklerck, R.: Context aware inactivity recognition for visual fall detection. In: Proc. IEEE Pervasive Health Conference and Workshops, pp. 1–4 (2006)
8. Kepski, M., Kwolek, B., Austvoll, I.: Fuzzy inference-based reliable fall detection using kinect and accelerometer. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) *ICAISC 2012, Part I. LNCS*, vol. 7267, pp. 266–273. Springer, Heidelberg (2012)

9. Kepski, M., Kwolek, B.: Fall detection on embedded platform using kinect and wireless accelerometer. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012, Part II. LNCS, vol. 7383, pp. 407–414. Springer, Heidelberg (2012)
10. Labayrade, R., Aubert, D., Tarel, J.P.: Real time obstacle detection in stereovision on non flat road geometry through “v-disparity” representation. In: IEEE Intelligent Vehicle Symposium 2002, vol. 2, pp. 646–651 (June 2002)
11. Marshall, S.W., Runyan, C.W., Yang, J., Coyne-Beasley, T., Waller, A.E., Johnson, R.M., Perkis, D.: Prevalence of selected risk and protective factors for falls in the home. *American Journal of Preventive Medicine* 8(1), 95–101 (2005)
12. Mastorakis, G., Makris, D.: Fall detection system using Kinect’s infrared sensor. *Journal of Real-Time Image Processing*, 1–12 (2012)
13. Miaou, S.G., Sung, P.H., Huang, C.Y.: A customized human fall detection system using omni-camera images and personal information. *Distributed Diagnosis and Home Healthcare*, 39–42 (2006)
14. Mubashir, M., Shao, L., Seed, L.: A survey on fall detection: Principles and approaches. *Neurocomputing* 100, 144–152 (2013); special issue: Behaviours in video
15. Noury, N., Fleury, A., Rumeau, P., Bourke, A., Laighin, G., Rialle, V., Lundy, J.: Fall detection - principles and methods. In: Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society, pp. 1663–1666 (2007)
16. Noury, N., Rumeau, P., Bourke, A., Olaighin, G., Lundy, J.: A proposal for the classification and evaluation of fall detectors. *IRBM* 29(6), 340–349 (2008)
17. Parra-Dominguez, G., Taati, B., Mihailidis, A.: 3d human motion analysis to detect abnormal events on stairs. In: Second Int. Conf. on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), pp. 97–103 (October 2012)
18. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Monocular 3D head tracking to detect falls of elderly people. In: Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society, pp. 6384–6387 (2006)
19. Tzeng, H.W., Chen, M.Y., Chen, J.Y.: Design of fall detection system with floor pressure and infrared image. In: Int. Conf. on System Science and Engineering, pp. 131–135 (2010)
20. Yu, X.: Approaches and principles of fall detection for elderly and patient. In: 10th Int. Conf. on e-health Networking, Applications and Services, pp. 42–47 (2008)

Evaluation of Various Techniques for SQL Injection Attack Detection

Michał Choraś and Rafał Kozik

Abstract. The network technology has evolved significantly recently. The growing use of cloud services, increased number of users, novel mobile operating systems and changes in network infrastructures that connect devices make novel challenges for cyber security. In order to counter arising threats, network security mechanisms and protection schemes also evolve and use sophisticated sensors and methods. In our previous work [27] we have introduced an innovative evolutionary algorithm for modeling genuine SQL queries generated by web-application. In [28] we have investigated how the proposed algorithm can be combined together with other Off-The-Shelf solutions (like SNORT and SCALP tools) in order to increase the detection ratio of injection attacks. In this paper we have significantly extended our test suite. First of all, we have compared our method with new efficient solutions for injection attack detection. We have also deeply discussed the drawbacks and benefits of these solutions. We have also explained how the correlation techniques can be adapted in order to overcome these drawbacks without loosing high effectiveness.

1 Introduction

Recently there is an increasing number of security incidents reported all over the world. In different parts of the world cybercriminals launch different malicious programs, their attacks have different priorities and they use different tricks. The analyses of the current situation is presented by Kaspersky Lab in Fig.1. Also the national CERTs (e.g. CERT Poland [1]) report that number of attacks in 2011 has increased

Michał Choraś
ITTI Ltd., Poznań, Poland
e-mail: mchoras@itti.com.pl

Michał Choraś · Rafał Kozik
Institute of Telecommunications, UT&LS Bydgoszcz, Poland
e-mail: chorasm@utp.edu.pl

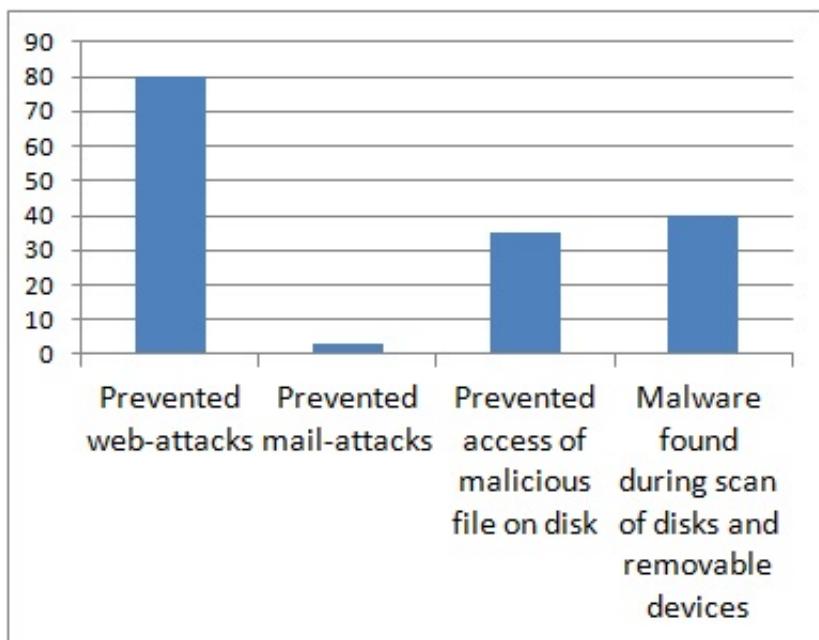


Fig. 1 The attack vectors in Western Europe and North America in the first half of 2012 [24]

significantly when compared to 2010. In annual reports they explain that most of network events submitted by automated feeds concern bot nets, spam, malicious URLs and Brute Force attacks.

The increased number of incidents is strongly related to the fact that recently there is also an increasing number of mobile devices users that form the population of connect-from-anywhere terminals that regularly test the traditional boundaries of network security.

According to ESET Global Threat Reports [25] [26], recently malware for Android devices is dominated by programs like Boxer, an SMS Trojan that covertly subscribes the victim to a premium rate SMS number. What is more ESET estimates that the growth of malicious code for Android has exponential nature, because the number of unique detection of malwares grew 17 times globally when compared to 2011.

In 2013, ESET also expects changes in the methods in which malicious programs are propagated. The approach using removable storage devices is constantly decreasing in popularity. Currently, preferred methods engage web servers that have been compromised by hackers in order to host malicious code. Users are redirected to infected URLs via emails. However, many of nowadays malwares like ZITMO (Zeus In The Mobile) do not aim at mobile device itself anymore but on gathering the information about the users and gaining the access to remote services like bank web services. This significantly expands cyber space network security perimeter.

Also as more and more cloud services and SaaS have been adapted by small and medium enterprises a big challenge for network security arises, since crucial for companies data started to be stored, maintained and transported by third party infrastructure where traditional points of inspection cannot be deployed. Storage in the cloud is one of such examples. Several cyber attack on cloud-base storage services (e.g. Dropbox) were reported. In case of Dropbox service the attack has not caused failure of the service since the attackers have accessed some accounts using stolen credentials, however the incident caused the company to improve its security. Other examples of companies affected by information leakage in recent years include LinkedIn, Yahoo and Formspring. Recently, also credit cards companies Visa and MasterCard have been affected by similar kind of cyber attacks that allowed the attackers to stole user data and to commit financial frauds. According to CISCO 2011 report [3] this trend is connected with the criminals that see the potential to get more return on their investment with cloud attacks, since they only need to 'hack one to hack them all'.

Other well known problems like attacks on the web applications to extract data or to distribute malicious code still remain unsolved. Cybercriminals continuously steal data and distribute their malicious code via legitimate web servers they have compromised. Also the emerging technologies such as HTML5 bring new cyber threats that web services providers have to deal with.

Among all attack targeted on web-servers the SQLIA (SQL Injection Attack) still remains one of the most important network threat which is ranked as one of the top threats in the OWASP list [8]. SQL injection and other similar exploits are the results of interfacing a scripting language by directly passing information through another language and are ultimately caused by insufficient input validation. SQL Injection Attacks (SQLIA) refer to a code-injection attacks category in which part of the user's input is treated as SQL code. Such code, if executed on the database, may change, erase, or expose sensitive data stored in the database.

One of the most significant examples of SQL Injection Attacks include:

- hacking the Royal Navy's website and recovering user names and passwords of the site's administrators (November 2010) [9];
- stealing information related to almost 100000 accounts of subscribers registered on ISP news and review site DSLReports.com (April 2011) [10];
- exploiting SQL injection vulnerabilities of approximately 500000 web pages (April-August 2008) [11].

2 Previous Work

In our previous work [27] we have introduced an innovative evolutionary algorithm for modeling genuine SQL queries generated by web-application. In [28] we have investigate how the proposed algorithm can be combined together with other Off-The-Shelf solutions (like SNORT and SCALP tools) in order to increase the detection ratio of injection attacks. In this paper we have extended our test suite. First of

all, we have compared our method with new efficient solutions for injection attack detection with. We have also deeply discussed the drawbacks and benefits of these solutions. We have also explained how the correlation techniques can be adapted in order to overcome these drawbacks without loosing high effectiveness. The set of tools used in this paper for detecting the SQL Injection attacks consists of both an algorithms proposed by authors and known (state of the art) solutions and tools. The state of the art tools evaluated in our tests are:

- Apache Scalp.

It is an analyzer of Apache server access log file. It is able to detect several types of attacks targeted on web application. The detection is a signature-based one. The signatures have form of regular expressions that are borrowed from PHP-IDS project.

- Snort.

It the most widely deployed IDS system that uses set of rules that are used for detecting web application attacks. However most of the available rules are intended to detect very specific type of attacks that usually exploit very specific web-based application vulnerabilities.

- ICD (Idealized Character Distribution [17]).

The method is similar to the one proposed by C.Kruegel in [17]. The proposed character distribution model for describing the genuine traffic generated to web application. The Idealized Character Distribution (ICD) is obtained during the training phase from perfectly normal requests send to web application. The IDC is calculated as mean value of all character distributions. During the detection phase the probability that the character distribution of a query is an actual sample drawn from its ICD is evaluated. For that purpose Chi-Square metric is used.

- PHP-IDS (PHP-Intrusion Detection System).

Is a simple to use, well structured, fast and state-of-the-art security layer for PHP based web applications. It is based on a set of approved and heavily tested filter rules. Each attack is given a numerical impact rating which makes it easy to decide what kind of action should follow the hacking attempt. This could range from simple logging to sending out an emergency mail to the development team, displaying a warning message for the attacker or even ending the userâŽs session [19].

- SQL-ADS based on the Genetic Algorithm (proposed by authors [27] and described in section 3.

Therefore, in this paper, we present real-time network data analysis mechanism and we prove its effectiveness for SQL Injection Attacks detection. The paper is structured as follows. In Section 3 we present our own solution for SQL Injection attempts detection based on the evolutionary algorithm. The experimental setup and results are provided in Sections 4 and 5. Conclusions are given thereafter.

3 Algorithm for Modeling the Normal Traffic

The details of algorithm form modeling the normal traffic were explained in our previous paper [27]. However, to make this paper self-containing in this section we have decided to explain general architecture and purpose of the proposed evolutionary algorithm.

The proposed method exploits genetic algorithm, where the individuals in the population explore the log file that is generated by the SQL database. Each individual aims at delivering an generic rule (which is a regular expression in form "SELECT [a-zA-Z]+ FROM patient WHERE name like [a-zA-Z]+") that will describe visited log line. It is important for the algorithm to have an set of genuine SQL queries during the learning phase. The algorithm is divided into the following steps:

- Initialization. Each individual and line from log file is assigned. Each newly selected individual is compared to the previously selected in order to avoid duplicates.
- Adaptation phase. Each individual explores the fixed number of lines in the log file (the number is predefined and adjusted to obtain reasonable processing time of this phase).
- Fitness evaluation. Each individual fitness is evaluated. The global population fitness as well as rule level of specificity are taken into consideration, because we want to obtain set of rules that describe the lines in the log file.
- Cross over. Randomly selected two individuals are crossed over using algorithm for string alignment. If the newly created rule is too specific or too general it is dropped in order to keep low false positives and false negatives.

The fitness function, that is used to evaluate each individual, takes into account the particular regular expression effectiveness (number of times it fires), the level of specificity of such rule and the overall effectiveness of the whole population. The fitness function is described by equation 1, where I indicates the particular individual regular expression, $E_{population}$ indicates the fitness of the whole population, E_f effectiveness of regular expression (number of times the rule fires), and E_s indicates the level of specificity (in order to avoid too short regular expressions like ".*"). The α , β , and γ are constants that normalize the overall score and balance the each coefficient importance.

$$E(I) = \alpha * E_{population} + \beta * E_f(I) + \gamma * E_s(I) \quad (1)$$

$$E_{population} = \sum_{I \in Population} E_f(I) \quad (2)$$

4 Experiments

In this section our evaluation methodology is described. The SQL Injection Attacks are conducted on php-based web service with state of the art tools for services penetration and SQL injection. The traffic generated by attacking tools are combined together with normal traffic (genuine queries) in order to estimate the effectiveness of the proposed methods. The genuine queries are both man-made and generated by web crawlers as well.

The web service used for penetration test is so called LAMP (Apache + MySQL + PHP) server with MySQL back-end. It is one of the most common worldwide used servers and therefore it was used for validation purposes. The server was deployed on Linux Ubuntu operation system. For penetration tests examples services developed in PHP scripts and shipped by default with the server are validated.

Attack injection methodology is based on the known SQL injection methods, namely: boolean-based blind, time-based blind, error-based, UNION query and stacked queries. For that purpose sqlmap tool is used. It is an open source penetration and testing tool that allows the user to automate the process of validating the tested services against the SQL injection flaws.

In order to avoid double-counting the same attack patterns during the evaluation process, we decided to gather first the malicious SQL queries generated by sqlmap (several hundreds of different injection trials). After that genuine traffic (generated by crawlers and during the normal web service usage) is gathered. Such prepared data is used during the evaluation test that results are presented in section 5.

The conducted experiments were aimed at estimating the effectiveness of different tools commonly used for injection attack detection. Namely these are:

- PHP-IDS (PHP GET and POST arrays),
- Apache-Scalp (HTTP access log),
- Snort (HTTP packet content),
- ICD (HTTP access log),
- proposed SQL-ADS (SQL DB log).

It must be noticed that PHP-IDS, Apache-Scalp as well as Snort tools do not require any learning phase, since the signatures of anomaly (having symptoms of SQL Injection) SQL queries and malicious HTTP request are provided together with thes tools. The signatures are developed by security experts in form of regular expressions. The ICD and SQL-ADS require dedicated learning phase and focus only on genuine HTTP and SQL queries. Method used for evaluation engages the classic 10-fold algorithm.

First experiment was intended to evaluate each method for SQL injection attack separately. All other of the experiments (from 2 to 7) were aimed at investigating whenever combining above methods together can additionally improve overall effectiveness of injection attack detection. Moreover, some of the presented scenarios were also conducted to emphasize the added value of propose SQL-ADS algorithm. For evaluation purposes a 10-fold approach is used. The information obtained from SQL-ADS, PHP-IDS, SCALP, ICD and SNORT is used to build classifier for attack

detection. In order to conduct the experiments WEKA [29] tool is used. Among the all classifiers provided by WEKA, following ones were reported to be statistically significant and to have satisfactory performance (better than each of the tool used separately):

- PART

The PART classifier uses separate-and-conquer and builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule. More details can be found in [18].

- NB (Naive Bayes)

Probabilistic Naive Bayes classifier based on applying Bayes theorem with strong (naive) independence assumption [20].

- REPTree

Fast decision tree learner. Builds a decision tree using information gain and prunes it using reduced-error pruning (with backfitting) [21].

- J48

Classifier that generates a pruned C4.5 decision tree [22].

- RIDOR

The implementation of a RIpple-DOwn Rule learner that performs a tree-like expansion of exceptions. The exceptions are a set of rules that predict classes other than the default [23].

- ADABOOST (Adaptive Boosting)

It is a meta-algorithm for machine learning and can be used in conjunction with many other algorithms to improve their performance.

In Table 1 configuration for different experiments are shown. We have investigated different configuration of scenarios for effectiveness evaluation. These scenarios try to reflect some deployment issues connected with mentioned above injection attack detection tools. For example PHP-IDS requires changes in HTTP server configuration in order to be operational. Moreover, PHP-IDS can only be deployed when the HTTP server supports PHP technology. When it comes to Apache-Scalp tool it capable of analyzing a HTTP-GET requests by default. Therefore, to increase detection effectiveness it is required to modify the server configuration.

Table 1 SQLIA detection sensors used in different scenarios (from 2 to 7). These address some deployment issues connected with injection attack detection tools.

	SQL-ADS	SNORT	ICD	SCALP	PHP-IDS
Experiment 2			x	x	
Experiment 3			x	x	x
Experiment 4	x		x	x	
Experiment 5	x		x	x	x
Experiment 6		x	x	x	x
Experiment 7	x	x	x	x	x

5 Results

As it is shown in Table 2, for the first experiment, the PHP-IDS algorithm slightly outperforms other approaches. When it comes to modelling the genuine queries proposed SQL-ADS is almost as good as PHP-IDS but for queries having the symptoms of attack, the SQL-ADS is about 10% worse when compared with ICD, which has the best performance of attacks detection.

Table 2 Effectiveness of injection attack detection (shown separately for genuine and malicious requests)

	SQL-ADS	SNORT	ICD	SCALP	PHP-IDS
Attack	87,8%	66,3%	97,9%	50,9%	93.5%
Genuine	97,7%	80,5%	94,5%	96,1%	98,1%
Weighted Avg.	96,2%	78,3%	95,0%	89,0%	98,1%

Table 3 Detectors used in different scenarios

	NaiveBayes	PART	Ridor	J48	REPTree	AdaBoost
Experiment 2	96.73	96.84	96.38	96.87	96.83	96.03
Experiment 3	97.54	97.06	96.99	97.02	97.12	96.07
Experiment 4	96.91	99.00	98.93	99.10	98.97	98.80
Experiment 5	98.35	99.03	99.02	99.12	98.99	98.89
Experiment 6	98.83	99.24	99.27	99.24	99.26	99.30
Experiment 7	99.08	99.51	99.40	99.54	99.37	99.67

The effectiveness for different configuration setups (experiments from 2 to 7) is shown in Table 3. It can be noticed that the highest effectiveness is reported for experiment 7, when all detectors are used. The effectiveness is increased by 1.5% when compared with PHP-ID and is reported to be 99.67%.

The experiments 2 and 4 have the no impact on server configuration. All detectors used in these scenarios (ICD, SCALP and SQL-ADS) are transparent for web servers since they process log files generated by HTTP and DB daemons. It can be noticed that when the proposed SQL-ADS is added the effectiveness increases by more than 2%. Without SQL-ADS the performance slightly worse but still better than each detectors individual effectiveness (which is 95% for ICD). When experiments 4 and 5 are compared it can be noticed that PHP-IDS does not increase the effectiveness which is 99.1%. Therefore, SQL-ADS, ICD and SCALP can be a good alternative when it is impossible to deploy PHP-IDS.

6 Conclusions

In this paper we have extended our previous work presented in [27] and [28]. The main contribution is an evaluation presented in results section. We have compared our method with new efficient solutions for injection attack detection. We have also deeply discussed the drawbacks and benefits of these solutions and explained how the correlation techniques can be adapted in order to overcome drawbacks without loosing high effectiveness. Our experiments show that combining several weak injection attack detectors and engaging the machine learning techniques can lead to overall effectiveness improvement. Our test suite has also addressed configuration issues, since some of the investigated solutions require significant deployment effort. In such cases presented results showed that effectiveness of several weak methods combined together is comparable to commercial tools (e.g. PHP-IDS).

Acknowledgements. This work was partially supported by Applied Research Programme (PBS) of the National Centre for Research and Development (NCBR) funds allocated for the Research Project number PBS1/A3/14/2012 (SECOR)).

References

1. CERT Polska Annual Report (2011),
<http://www.cert.pl/PDF/ReportCP2011.pdf>
2. SOPHOS homepage, <http://www.sophos.com>
3. Cisco Annual Report (2011)
4. Choraś, M., Kozik, R., Piotrowski, R., Brzostek, J., Hołubowicz, W.: Network Events Correlation for Federated Networks Protection System. In: Abramowicz, W., Llorente, I.M., Surridge, M., Zisman, A., Vayssièvre, J. (eds.) ServiceWave 2011. LNCS, vol. 6994, pp. 100–111. Springer, Heidelberg (2011)
5. Rao, T.K., Kum, G.Y., Reddy, E.K., Sharma, M.: Major Issues of Web Applications: A Case Study of SQL Injection. Journal of Current Computer Science and Technology 2(1), 16–20 (2012)
6. Halfond, W., Orso, A.: AMNESIA: Analysis and Monitoring for Neutralizing SQL-Injection Attacks. In: Proceedings of the 20th IEEEACM International Conference on Automated Software Engineering (2005)
7. <https://paulsparrows.wordpress.com/2011-cyber-attacks-timeline-master-index/>
8. OWASP Top 10 2010, The Ten Most Critical Web Application Security Risks (2010)
9. Royal Navy Website Attacked by Romanian Hacker (2008),
<http://www.bbc.co.uk/news/technology-11711478>
10. Mills, E.: DSL Reports Says Member Information Stolen (2011)
11. Keizer, G.: Huge Web Hack Attack Infects 500,000 pages (2008)
12. Tajpour, A., Jor Jor Zade Shooshtari, M.: Evaluation of SQL Injection Detection and Prevention Techniques. In: CICSyN 2010, Second International Conference on Computational Intelligence, Communication Systems and Networks (2010)

13. Amirtahmasebi, K., Jalalinia, S.R., Khadem, S.: A Survey of SQL Injection Defense Mechanisms. In: ICITST International Conference for Internet Technology and Secured Transactions (2009)
14. Elia, I.A., Fonseca, J., Vieira, M.: Comparing SQL Injection Detection Tools Using Attack Injection: An Experimental Study. In: 2010 IEEE 21st International Symposium on Software Reliability Engineering (2010)
15. Needleman, S.B., Wunsch, C.D.: A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology* (1970)
16. Conrad, E.: Detecting Spam with Genetic Regular Expressions. SANS Institute InfoSec Reading Room (2007)
17. Kruegel, C., Toth, T., Kirda, E.: Service specific anomaly detection for network intrusion detection. In: Proc. of ACM Symposium on Applied Computing, pp. 201–208 (2002)
18. Frank, E., Witten, I.H.: Generating Accurate Rule Sets Without Global Optimization. In: Fifteenth International Conference on Machine Learning, pp. 144–151 (1998)
19. PHP-IDS project homepage, <https://phpids.org/>
20. John, G.H., Langley, P.: Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, pp. 338–345 (1995)
21. Weka REPTree reference manual, <http://www.dbs.informatik.uni-muenchen.de/zimek/diplomathesis/implementations/EHNDs/doc/weka/classifiers/trees/REPTree.html>
22. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo (1993)
23. RIDOR classifier, <http://weka.sourceforge.net/doc/weka/classifiers/rules/Ridor.html>
24. Kaspersky Lab. Security report, <http://www.securelist.com/en/analysis/204792244/The-geography-of-cybercrime-Western-Europe-and-North-America>
25. ESET annual report, <http://go.eset.com/us/resources/white-papers/Trends-for-2013-preview.pdf>
26. ESET threat report (December 2012), <http://go.eset.com/us/resources/threat-trends/Global-Threat-Trends-November-2012.pdf>
27. Choraś, M., Kozik, R., Puchalski, D., Hołubowicz, W.: Correlation Approach for SQL Injection Attacks Detection. In: Herrero, Á., Snášel, V., Abraham, A., Zelinka, I., Baruque, B., Quintián, H., Calvo, J.L., Sedano, J., Corchado, E., et al. (eds.) Int. Joint Conf. CISIS'12-ICEUTE'12-SOCO'12. AISC, vol. 189, pp. 177–185. Springer, Heidelberg (2013)
28. Choraś, M., Kozik, R.: Real-Time Analysis of Non-stationary and Complex Network Related Data for Injection Attempts Detection. In: Proc. of WSC 17 Online Conference on Soft Computing in Industrial Applications (2012)
29. WEKA 3 Data mining tool homepage, <http://www.cs.waikato.ac.nz/ml/weka/>
30. Ficco, M., Coppolino, L., Romano, L.: A Weight-Based Symptom Correlation Approach to SQL Injection Attacks. In: Fourth Latin-American Symposium on Dependable Computing, LADC 2009, September 1-4, pp. 9–16 (2009)

Task Allocation in Distributed Mesh-Connected Machine Learning System: Simplified Busy List Algorithm with Q-Learning Based Queuing

Agnieszka Majkowska, Dawid Zydek, and Leszek Koszalka

Abstract. In the era where organizations gather and process more and more data, Machine Learning (ML) techniques become increasingly important. Considering “Big Data,” ML usually involves intensive data processing and high-performance computing. To meet the growing requirements, efficient distributed and parallel systems are key factors. In this paper, we consider mesh-based distributed system and task allocation methods in the system. We focus especially on the impact of intelligent queuing in task allocation algorithms. A new SBL algorithm with Q-Learning queuing is presented. In addition, the new SBL technique is compared to other well-known allocation schemes, which are discussed as well. The comparison is made using an implemented experimentation system and simulation results are presented. The results confirm that SBL algorithm and the queuing system deliver good performance characteristic.

1 Introduction

Machine Learning (ML) techniques are used commonly in many areas like e.g. medical applications; image, pattern, or speech recognition; decision-making systems; or “Big Data” analysis and processing. Due to wide use and huge potential, ML is a subject of significant research interest. However, advancement and complexity of ML algorithms, together with very large and heterogeneous data sets processed

Agnieszka Majkowska · Leszek Koszalka
Department of Systems and Computer Networks
Wrocław University of Technology, Poland

e-mail: aga.majkowska@gmail.com, leszek.koszalka@pwr.wroc.pl

Dawid Zydek
Department of Electrical Engineering
Idaho State University, USA
e-mail: zydedawi@isu.edu

by them, cause increasing demand on computation power. It makes single-server systems not sufficient enough; and multiple, very often distributed servers or supercomputers working in parallel are needed [1, 2, 5].

Distributed parallel systems, where Processing Elements (PEs) are organized in a mesh topology are substantial solutions for demanding computations, e.g. ML systems [8]. Mesh network consists of several, identical nodes with the same processing power, and nodes have a possibility to communicate among each other. In this kind of structure, an incoming job (task), part of ML algorithm, requests a specific number of PEs to be completed. To decrease the communication cost, the PEs used are adjacent. Besides energy constraints in such a powerful structure, the main challenge is to use its maximum processing power, that is basically unreachable. It is related to the task allocation problem. The processor allocation is the process of selecting and allocating available PEs to an incoming job. The optimal allocation is when the highest amount of PEs is used in parallel and when the tasks are computed in a short time [13–15].

In this paper, a new allocation algorithm, Simplified Busy List (SBL), is introduced. The technique is an improved but simplified version of well-known Busy List scheme [7], where the scoring system is changed. In addition, a new way of queuing tasks using reinforcement learning (basic Q-Learning algorithm) is presented.

The rest of this paper is organized as follows. In Section 2, basic terminology used in the paper and problem statement are presented. Section 3 contains related work and description of the new SBL algorithm. In Section 4, Q-Learning algorithm and its sample use in task allocation problem are discussed. Experimentation system and results are in Section 5. The paper is concluded in Section 6.

2 Task Allocation Problem

Basic Terminology

The mesh is a set of nodes (PEs) connected among each other. It is represented as a matrix $M = (w, h)$ with specified size w (width) and h (height). *Node* is an element of the mesh, labeled with (i, j) , where i is a column number and j is a row number. The node can be free or busy. *Submesh* is a rectangular collection of nodes, it is represented as a set $S = (i, j, x, y)$, where (i, j) is a starting node, x is number of columns, and y is number of rows. Submesh can be *free* (all its nodes are free) or *busy* (at least one of its nodes is busy). *Task* is a rectangular set of nodes representing PEs. The number of nodes describes an amount of PEs needed to complete a task. A task is represented by three parameters (x, y, t) ; where x is width of the task, y is height of the task, and t is time needed to complete the task. Tasks are buffered in a *queue* while waiting for allocation. To allocate a task, it is required to find free submesh that is big enough to accommodate all nodes. *Allocation* of a task is a creation of busy submesh that has the same size as the task. *Deallocation* is reduced to changing the status of the busy nodes to free, when all nodes finished computations and a task is completed. *Fragmentation* is a ratio of busy nodes to the total num-

ber of nodes in the mesh. *Internal fragmentation* may occur after task allocation if there are more nodes in a submesh than needed. *External fragmentation* may happen when there is enough nodes to allocate a task but the nodes are not adjacent thus the task cannot be allocated. *Blockade* is a moment when it is not possible to allocate a waiting task. The system has to wait until some already allocated and big enough tasks will be completed and free from the mesh. *Task allocation algorithm* is an algorithm that leads to allocation of PEs to tasks waiting in a queue. Optimal allocation means that the performance of the mesh is as high as possible.

Problem Statement

GIVEN (1) The mesh with specified size $M = (w, h)$; (2) A list of tasks $T = T_1, T_2, \dots, T_N$ with well-known characteristics $T_p = (x_p, y_p, t_p)$,

TO FIND Sequence of tasks to be allocated,

SUCH THAT Parameters, i.e. average fragmentation, completion time, and average allocation time; reach required level,

SUBJECT TO CONSTRAINTS (1) Once task is allocated it cannot be moved to another free submesh; (2) Once task is allocated it has to stay in the mesh until it is completed; (3) In order to allocate a task, a free submesh, not smaller than the task being allocated, must exist; (4) If allocation is not possible, the system waits; (5) The total number of nodes in the mesh is not smaller than the total number of nodes needed to complete a task.

Tasks intend to be allocated in the mesh are waiting in a queue. When a task from the queue is ready for allocation and there is a large enough free submesh, allocation can be performed. After execution of the task, it can be deallocated. An example of the mesh with allocated tasks can be found in [13].

The problem is solved when all the tasks from queue are executed. The moment of solving allocation problem is when the last task quits the mesh. There are two aspects of the allocation problem. The first one is a task scheduling in a queue, another one is a task allocation in the mesh. In this paper, both aspects are considered.

Three measures of algorithm's performance are employed:

- *Average fragmentation of the mesh*, which is a ratio between the number of free nodes and the total number of nodes in the mesh. It is measured each time a blockade occurs and it is an average taken out of all considered measures. It is desirable to minimize this factor.
- *Completion time*, which is a moment when the last task quits the mesh. The considered system is implemented using time-slices and any event (allocation/deallocation) is triggered by this unit. The less time-slices needed to finish computations, the highest performance of the mesh is obtained.
- *Average allocation time*, which describes time in milliseconds used by algorithm to find a free submesh (to allocate a task).

3 Allocation Algorithms

There are two types of allocation schemes in mesh systems: contiguous and non-contiguous [13]. In this paper we focus only on contiguous allocation algorithms.

Related Work

Frame Sliding (FS) algorithm [3] looks for a place to allocate a task frame by frame. Frame is in size of the task. FS creates a set of busy nodes called coverage set. The coverage set is then scanned to find a frame with free nodes. FS causes high external fragmentation of the mesh and it might miss some free submeshes.

Adaptive Scan (AS) [4] is similar to FS, but it scans the coverage set node by node. In addition, it allows rotating a task if it is not possible to allocate it in its basic orientation. AS needs much more time to allocate task than FS, but it is more precise.

Quick Allocation (QA) [11] is a development of FS and AS strategies, where individual checking of the nodes is replaced by testing only each row in the mesh. QA is fast but not efficient for large meshes. **First Fit (FF)** [12] creates an array of nodes called busy array. It represents an occupancy status of the nodes. Based on this array, FF forms an array called coverage array that contains information about nodes allowing successful allocation of the task. FF selects the first found node as a starting node. **Best Fit (BF)** [12] is similar to FF, the only difference is the way of choosing node – it selects the best one to minimize external fragmentation. **Stack-Based Allocation (SBA)** [10] finds a free submesh very fast by reducing the search space significantly. It is done by using simple coordinate calculations and spatial subtraction. All these operations are implemented proficiently using stack. There are other variances of this algorithm, e.g. **Better Fit SBA (BFSBA)** [6] that chooses the smallest candidate from the stack, Sorted SBA [6], or Improved SBA [13]. **Leapfrog (LF)** algorithm [9] creates R-Array (Run-length Array) that represents nodes located in the mesh. Every element of this array represents a node of the mesh and its value expresses the length of the free or occupied run counted from this node. To search for suitable submeshes, LF scans elements in R-Array. There is also **Leapfrog Best Fit (LBF)**, where the algorithm looks for the best place to allocate a task (like BF). Implementation of R-Array shortens the process of scanning the mesh. **Busy List (BL)** [7] uses an idea of tightening tasks in the mesh to reduce the external fragmentation. The algorithm works in two stages: firstly, it creates a list of candidates and then it selects the best candidate. To create candidates, it checks all tasks in the busy list and all the adjacent nodes in the list of free submeshes. In addition, it adds four corners of the mesh and creates list of candidates. Then, the candidates are evaluated by checking the occupancy status of nodes adjacent to the allocated task (using the specific candidate). The higher number, the better. BL reduces external fragmentation of the mesh but the scoring system increases execution time of BL.

New Algorithm – Simplified Busy List (SBL)

Simplified Busy List (SBL) is based on BL technique. The scoring system is changed that shortens the allocation time. It is done by improving the second stage of BL

scheme. To reduce candidate's evaluation time, SBL checks only four corners instead of checking all potential submeshes (it is $2*x + 2*y$). Rules in scoring system:

- Candidate gets 1 point for every busy node in the corner.
- If candidate is located at the edge of the mesh, it gets 2 points.
- If one corner of the candidate is surrounded by busy nodes, it gets 2 points.
- If the difference between computation time of the task being allocated and completion time on candidate's nodes adjacent to the corners is not bigger than 20% of maximum computation time of tasks, then the candidate gets 1 point for each adjacent node meeting the criteria.

A candidate with the highest score is selected. The scoring system is shown in Fig. 1. SBL technique “thinks” forward and groups tasks allocated in the mesh by their computation time, in order to create larger free submeshes. Structure of SBL algorithm is presented in Fig. 2.

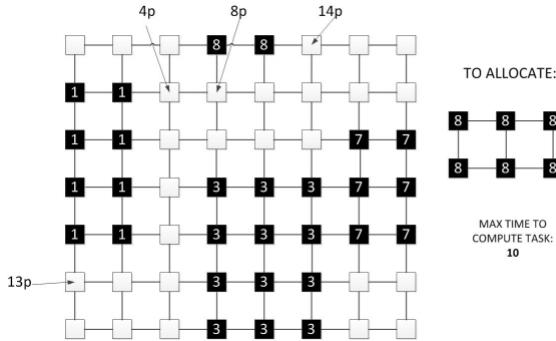


Fig. 1 Scoring system in SBL algorithm

4 Reinforcement Learning

Q-Learning Algorithm

Q-Learning agent learns Q-function, giving the expected utility of taking a given action in a given state. It strives to maximize this function. Q-function has an array-based representation. Therefore value $Q(x, a)$ for a state x and an action a is represented as exactly one element of an array, which represents expected value of sum of upcoming rewards, which system would get if it starts in state x and makes an action a .

The Q-function is updated after an action is made using following expression:

$$Q(x_t, a_t) \leftarrow -(Q(x_t, a_t) + \beta(r_t + \gamma \max_a Q_t(x_{t+1}, a) - Q_t(x_t, a_t))) \quad (1)$$

The factor $\beta[0, 1]$ is a learning factor and it specifies a greatness of modification. The factor γ is a discount factor $[0, 1]$ and it describes an importance of award. r_t is a reward of taking an action a_t in state x_t .

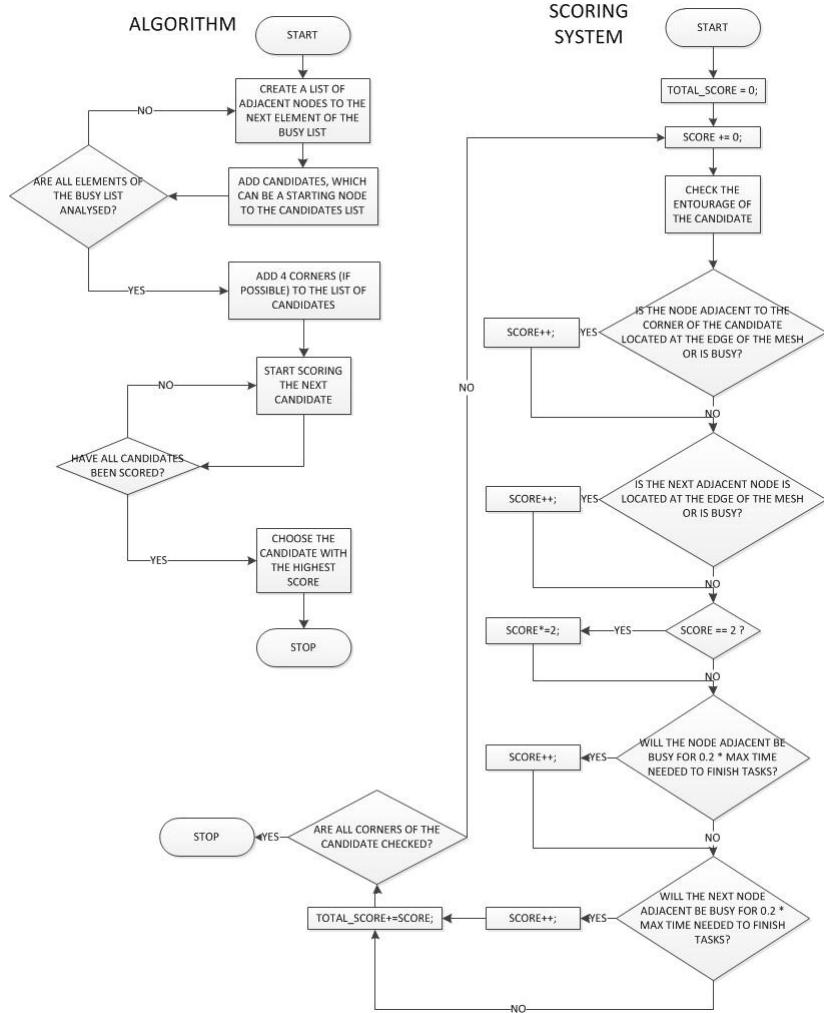


Fig. 2 Flowchart of SBL algorithm

The agent chooses an action according to a strategy π . In our implementation, the Boltzman action selection is used:

$$\pi(x, a^*) = \frac{\exp(\frac{Q(x, a^*)}{T})}{\sum_a \exp(\frac{Q(x, a)}{T})}, \quad (2)$$

where T is a temperature specifying degree of randomness. The higher temperature, the more random is selection of action.

Use of Q-Learning Algorithm in Task Allocation

Q-Learning algorithm is used to create a queue of tasks that allows their optimal allocation. State is a permutation of elements in the queue. A set of states is a set of all possible permutations. Action is an exchange between two elements. A set of actions is a collection of sets of two-factor without repetition. The reward is based on the completion time, which is obtained after an allocation of the specific permutation by using chosen allocation algorithm.

After obtaining the reinforcement, a Q-value can be updated. To find an optimal queue, actions and permutations with the highest Q-value are selected. Since an array-based representation of Q-function is memory consuming, the queue is divided and sorted part by part. Maximum ten elements in one portion are allowed. Therefore for 100 tasks to sort, they are divided on 10 parts, sorted using Q-Learning, and then these 10 parts are sorted again.

5 Experimentation System and Results

An experimentation system to test allocation algorithms was design and developed. The system is represented as an input-output system similarly like in [14]. An input is represented by mesh and task parameters, task allocation algorithms (FS, AS, FF, BF, SBA, BFSBA, BL, LF, LBF, QA, and SBL), types of queues (First-In-First-Out (FIFO), Longest Job First (LJF), and based on Q-Learning (Q)), and parameters of Q-Learning algorithm. An output contains completion time, average fragmentation, and average allocation time. The experimentation system was created in C# using Microsoft Windows with Microsoft .NET Framework 4.0 installed.

Experiment 1: Allocation Algorithms in Function of the Mesh Size

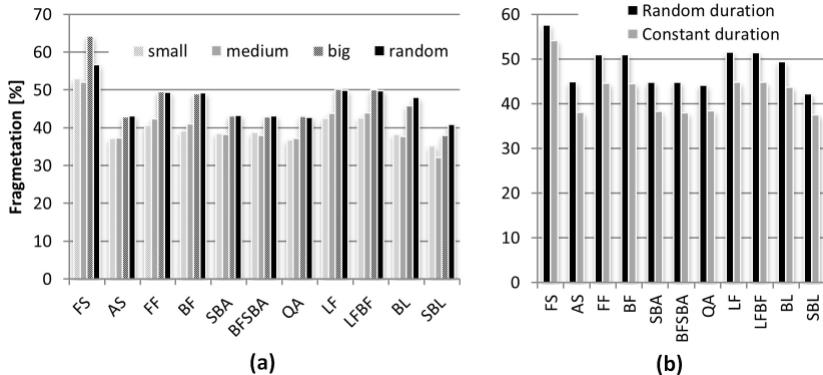
The meshes with size between 25 and 100 nodes were considered. A FIFO queue contained 100 tasks varying in size. The results are presented in Table 1. Task allocation time for the largest mesh is the highest for FF, BF, and QA algorithms; and they are the most sensitive to growth of the mesh. Similar problems are not observed for algorithms SBA, LF, and FS; and they are characterized by short task allocation time despite the mesh size. There is a difference between allocation times for algorithms and their Best Fit versions, e.g. LF and LBF techniques. The versions that decreasing external fragmentation have longer allocation time due to considering all possible places in the mesh and selecting the best candidate. The presented SBL algorithm is characterized by better allocation time in comparison to its original version, BL. The fastest algorithm is SBA thanks to eliminating candidates.

Experiment 2: Allocation Algorithms in Function of Varying Size of Tasks

Four groups of tasks were created: (i) small (max 15% of the mesh surface), (ii) medium (2% – 30%), (iii) big (10% – 50%), (iv) random. A 20×20 mesh was considered with 100 tasks in a FIFO queue. The results are presented in Fig. 1a.

Table 1 Comparison of task allocation time based on different mesh size

Size of the Mesh	Task Allocation Time [ms]			
	25 × 25	50 × 50	75 × 75	100 × 100
FS	0.01	0.01	0.10	0.30
AS	0.01	0.22	1.30	3.56
FF	0.55	10.65	51.76	145.05
BF	0.58	10.88	52.77	148.30
SBA	0.01	0.01	0.02	0.16
BFSBA	0.01	0.01	0.02	0.16
QA	0.38	4.26	14.54	31.69
LF	0.01	0.02	0.21	0.84
LBF	0.01	0.77	3.24	6.87
BL	0.01	0.41	1.59	3.37
SBL	0.01	0.25	1.28	2.83

**Fig. 1** Fragmentation in the mesh considering: (a) different size of tasks; (b) different duration of tasks

It may be observed that all algorithms achieve the worst results when random tasks are being allocated. The exception is FS scheme where results for random tasks are better than for big ones. It is due to structure of FS technique that uses a frame and for big tasks it can miss some free submeshes.

Experiment 3: Allocation Algorithms in Function of Varying Task Duration

Two FIFO queues with 100 tasks were created. One contained tasks with the same duration. Another one had tasks with random duration. A 20×20 mesh was considered. The results are shown in Fig. 1b.

Lower fragmentation is achieved for tasks with the same duration, in comparison to tasks with random duration, where fragmentation was higher. It is because allocation of one and deallocation of the next task are performed in the same time.

Experiment 4: Allocation Algorithms in Function of Different Queue Organization

Three types of queues with 100 tasks were used: FIFO, LJF, and Q. A 10×10 mesh was considered. Q-Learning algorithm was implemented with parameters: (i) number of iterations: 1000; (ii) discount factor: 0.8; (iii) learning factor: 0.2; (iv) temperature: 1000. The results are shown in Fig. 2.

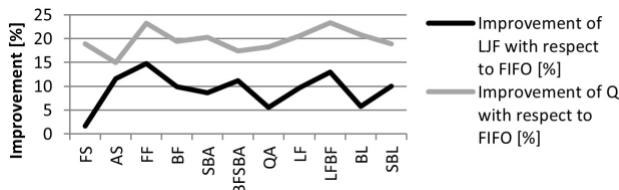


Fig. 2 Fragmentation improvement in respect to queuing type

It can be observed that each algorithm delivers better results when some kind of sorting is used before allocation process. Even LJF queue, which is characterized by simple implementation, improves performance of algorithms by around 10%. Moreover, using sorting based on Q-Learning algorithm, a 20% fragmentation improvement is noted.

6 Conclusions

In this paper we have introduced a new task allocation algorithm SBL, where Q-Learning algorithm was proposed and employed in the presented technique. In order to compare our scheme to other solutions, an experimentation environment was developed and several experiments were conducted. Results of the experiments are presented and discussed in the paper. The results have shown that among all described algorithms and for all considered sizes of tasks, our SBL algorithm delivers very good performance. Also, sorting tasks before allocation improves efficiency of all investigated techniques. Especially, Q-Learning based queue brings 20% improvement. The further research will focus on other ways of using machine learning in task allocation problem.

References

1. Arimond, A.: A distributed system for pattern recognition and machine learning. Master's thesis, TU Kaiserslautern and DFKI, Kaiserslautern (2010)
2. Borkar, V.R., et al.: Declarative systems for large-scale machine learning. IEEE Data Engineering Bulletin 32(2), 24–32 (2012)

3. Chuang, P.J., Tzeng, N.F.: An efficient submesh allocation strategy for mesh computer systems. In: 11th Inter. Conf. on Distr. Comp. Sys., pp. 256–263 (1991)
4. Ding, J., Bhuyan, L.N.: An adaptive submesh allocation strategy for two-dimensional mesh connected systems. In: Inter. Conf. on Para. Proc (ICPP 1993), vol. 2, pp. 193–200 (1993), doi:10.1109/ICPP.1993.39
5. Kamath, U., Kaers, J., Shehu, A., De Jong, K.A.: A spatial EA framework for parallelizing machine learning methods. In: Coello, C.A.C., Cutello, V., Deb, K., Forrest, S., Nicosia, G., Pavone, M. (eds.) PPSN 2012, Part I. LNCS, vol. 7491, pp. 206–215. Springer, Heidelberg (2012)
6. Koszalka, L., Kubiak, M., Pozniak-Koszalka, I.: Comparison of SBA – family task allocation algorithms for mesh structured networks. In: Min, G., Di Martino, B., Yang, L.T., Guo, M., Rünger, G. (eds.) ISPA Workshops 2006. LNCS, vol. 4331, pp. 21–30. Springer, Heidelberg (2006)
7. Tarnawski, M., Tarasiewicz, A.: Dynamiczna alokacja zadan w sieciach MESH. In: Students' Science Conference, Wroclaw University of Technology (2010)
8. Upadhyaya, S.R.: Parallel approaches to machine learning—A comprehensive survey. *J. of Para. and Distr. Comp.* 73(3), 284–292 (2013)
9. Wu, F., et al.: Processor allocation in mesh multiprocessors using the leapfrog method. *IEEE Trans. on Para. and Distr. Sys.* 14(3), 276–289 (2003)
10. Yoo, B., Das, C.: A fast and efficient processor allocation scheme for mesh-connected multicompilers. *IEEE Trans. on Comp.* 51(1), 46–60 (2002)
11. Yoo, S.M., Youn, H.Y., Shirazi, B.: An efficient task allocation scheme for 2D mesh architectures. *IEEE Trans. on Para. and Distr. Sys.* 8(9), 934–942 (1997)
12. Zhu, Y.: Efficient processor allocation strategies for mesh-connected parallel computers. *J. of Para. and Distr. Comp.* 16(4), 328–337 (1992)
13. Zydek, D., Selvaraj, H.: Fast and efficient processor allocation algorithm for torus-based chip multiprocessors. *J. of Comp. & Elect. Eng.* 37(1), 91–105 (2011), doi:10.1016/j.compeleceng.2010.10.001
14. Zydek, D., et al.: Evaluation scheme for NoC-based CMP with integrated processor management system. *Inter. J. of Elect. and Telecommu.* 56(2), 157–168 (2010), doi:10.2478/v10177-010-0021-4
15. Zydek, D., et al.: Energy characteristic of a processor allocator and a network-on-chip. *Inter. J. of Applied Math. and Comp. Sci.* 21(2), 385–399 (2011), doi:10.2478/v10006-011-0029-7

Power Saving Algorithms for Mobile Networks Using Classifiers Ensemble

Rafal Lysiak and Marek Kurzynski

Abstract. The main objective of this paper is to tackle the energy consumption for cellular radio networks. The mobile telecommunications system are optimized for the maximum load. Therefore, in the low traffic moment, the system consume incredible amounts of energy, which is not used in any way. The solution, which we propose in this paper is based on automatic switching on and off the network elements, depending on the current state of the network and on the prediction of the next state. It is also shown, that with the predictions from the ensemble of classifiers, the energy consumption can be reduced dramatically and such approach is acting better than simply setting the threshold values. The biggest challenge is to maintain reliable service coverage and quality of service (QoS) in the specific cell in the network.

1 Introduction

The motivation for starting this research that was described in this paper was the fact, that the large amounts of energy are absorbed by the mobile telecommunications network. According to studies, base stations for mobile networks consume approximately 60 billion kWh per year [1], and this consumption is still growing. Due to the appearance of new technologies, such as Universal Mobile Telecommunications System (UMTS), High Speed Packet Access (HSPA), and Long Term Evolution (LTE) and due to the continuous increasing the number of subscribers and the amount of transmitted data, operators add new elements to the network. According to the research [1], the electricity consumption of Vodafone network growth from about 1.8 TWh per year to almost 3 TWh per year in only 2 years (from 2003 to 2005).

Rafal Lysiak · Marek Kurzynski
Wroclaw University of Technology Dept. of Systems and Computer Networks,
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
e-mail: {rafal.lysiak, marek.kurzynski}@pwr.wroc.pl

The another important issue, which follows directly from the size of the electricity that is consumed, is the global emission of CO₂ and global warming. It is known, that about 2% of global CO₂ emmision - which is exactly the same proportion as the aviatation industry produces - comes from the Information and Communications Technology (ICT) [2]. Furthermore, it is estimated that this number will double by 2020.

The issue of savings in the described area is currently very up to date at the international forum. Last year, the european project called "Greennets" has been ended [4] [3]. It consisted in power consumption and CO₂ footprint reduction in mobile networks by advanced automated network management approaches. The extensive cooperation among mobile vendors was established, which lead to the creation of mechanisms responsible for reducing the amount of absorbed energy and reduce CO₂ emissions.

There are many opportunities for reducing energy consumption in mobile networks. Starting with the improvement of network elements, i.e. increasing the power amplifier efficiency ending on the effective use of the same elements from the software point of view [5]. In this paper, a self learning algorithm is proposed, which main target is to match the network configuration to the current traffic. As it was already mentioned, currently, the mobile networks are designed for maximum expected throughput and are optimized for operation at full load. However, it is easy to note, that the traffic is not homogeneous and the changes are significant during the time. The identification of such situations and their prediction can lead to a reduction in energy consumption, but also to a more efficient use of the current network structure. Mobile network traffic is not only voice data but more often the data packet. Rather than adding new elements to the network, we can use those elements that already are in the network and won't be used in particular moment.

The paper is organized as follows. In the section 2, two different approaches to saving energy in the network have been described. In the next section 3, the problem, proposed algorithms and research environment have been described. In section 4, the results of two different experiments that have been carried out were presented. In the end, in section 5, the conclusions were presented and ideas for future research were mentioned.

2 Identify Energy Saving Opportunities

With reference to what has been written in the first section of this work, the goal of the presented algorithms is to match the network configuration to the current network traffic characteristics and to predict the next state. It will be possible turning on and off the network elements in such a way as to ensure an appropriate level of QoS, while reducing the number of operating equipment at the same time. Analysing traffic patterns, it could be noticed, that traffic has stochastic characteristic and is changing during the time (day and night), but also in terms of geographic location of the subscriber.

2.1 Temporal Traffic Changes

In order to show the variation over time, the actual data provided by the vendor were analyzed. Those data was recorded by the measurement tools installed in a real mobile network. In the Fig. 1 the traffic changes during a day time is presented. The characteristic pattern is visible. In the middle of the day, this traffic is growing and the closer to the end of the day the smaller the traffic is. The traffic at night is almost 600% smaller than during the day time. In addition, the chart also shows how the traffic depends on the day of the week. At the weekend the traffic is much smaller what has been highlighted with the circle on the same figure. Network capacity is designed to provide a high QoS for the end user. According to that statement, the mobile network must guarantee a low probability of call blocking and call dropping. This approach leads to the fact that the cells and networks are seldom under full load. Today, hardware, such power amplifiers and signal processing boards operate continuously with full power, even if the network traffic is minimal [6].

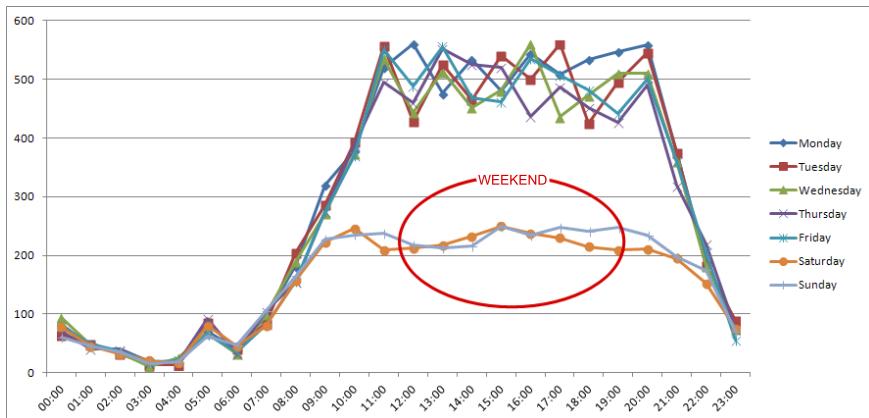


Fig. 1 Hourly Traffic Volume of CS Traffic [Erl]

Based on the Fig. 1, it is easy to discover that by about 20-30% of the day, traffic on the network is at a level much lower (about 65 %) than during the heavy load hours. The Fig. 1 has been prepared on the basis of the data obtained from the actual mobile network operator. Unfortunately, the exact location of the cell in which the readings were recorded can not be specified. Please contact the authors in order to obtain the access to these data.

2.2 Spatial Traffic Changes

As the name of "mobile" network suggests, we are dealing with mobile users. Analysing the geographical location of end-users, some patterns in daily or weekly

cycle could be noticed. In order to illustrate the situation, let us consider the base station, which is located on the border between urban and suburban land for big city. During the week, customers are moving in the morning towards the city, and then they return to the suburbs in the afternoon (after work). It is also known, that the type of traffic (voice or data) varies, depending on the geographical location.

One possibility using such a trend is already used in the Global System for Mobile Communications (GSM) and UMTS networks - the idea of micro and macro cells. In urban areas, the range of one cell is even just 500m, while in non-urban areas can cover an area with a radius of several kilometers. Because of this, an appropriate level of QoS can be provided, while decreasing variations in the traffic at the same time.

Another possibility is to modify the directivity of the antenna. The antenna should cover the area with big number of subscribers, instead of the area where subscribers aren't at the current moment.

In this paper, for the conducted research, changes that have been described in the subsection 2.1 were analyzed.

3 Problem, Methods and Research Environment Description

3.1 Problem Overview

In this paper results of simulations performed with respect to the subject of power savings are presented. The general problem was formulated as follows. According to the real traffic data, that was provided by the mobile vendor the traffic generator was implemented. The generator is used to perform the simulations that cover one cell with defined number of channels. Calls attempts arrive according to Poisson distribution with dynamic mean arrival rate taken from the real traffic data. The mean call holding time is exponentially distributed with fixed average value set to 100s. The presented traffic profile implies a scenario with 4 transceivers (TRXs). Each TRX has 8 channels available. 2 channels in the cell are reserved for signaling. Therefore, there are 30 channels available in the cell for voice transmissions.

One of the most important measurement that is collected in the following simulation is *blocking*. In the simulation scenario, *blocking* is considered as the situation where there are no free resources (no free channels) for new incoming call.

According to the Algorithm 1, the main issue of the performed simulation is to measure three important statistics when different approaches of turning on/off of the TRX was implemented. First value is *blocking*, which directly affects the subscribers' experience and thus the level of QoS. The second value is the average number of active TRXs. The main issue is to keep the energy consumption on the possible low level (small number of active TRXs) but still having a good QoS. The last measured parameter is the average time between turning on/off of the TRXs.

Algorithm 1. Pseudocode of the simulation

Require: \mathcal{T} - real traffic data set for the whole day;
 $time$ - time in the simulation for which the calls are generated; initial value is set to 0;
 end - this value represents the duration of the simulation; set to 21 days (1814400s);
 $step$ - by this value the $time$ is incremented in the simulation loop; initial value of $step$ is set to 0.1s;
 $Calls$ - an array that stores ongoing calls; initially empty table

- 1: **while** $time < end$ **do**
- 2: Check and removed finished calls from $Calls$
- 3: Check and start new calls; save the information in $Calls$
- 4: Check if the TRX should be turn on/off (different approaches are taken into consideration)
- 5: Collect statistics about Blocking, number of active TRXs, time between turning TRX on/off
- 6: $time \leftarrow time + step$
- 7: **end while**

It is assumed that the TRX is ready 30 seconds after the decision was made. Therefore, it is expected to maximize the time between switching.

As it was also mentioned in the pseudocode, the biggest challenge of the presented algorithm is to make the correct decision, if the TRX should be turn on or off. In this paper we use two different approaches. In the next subsections, the both methods are described.

3.2 Method 1 - Threshold Algorithm

The first method is very intuitive. The decision about switching the TRX off is based on the following algorithm:

The value 8 in the Algorithm 2 is required, because there have to remain T_{down} number of free channels to keep QoS at the good level. Otherwise, after the TRX would be powered off, there won't be free channels and the number of *blocking* calls would increase dramatically.

On the other hand, the decision about turning the TRX on is triggered immediately. The algorithm is very similar to the previous one. The threshold value T_{up} is set and when $N_{fc} \leq T_{up}$ is true, the TRX is powered up. In this case, the decision is made without any delay, because as it was written in the subsection 3.1, the time required for turning the TRX on is set to 30 seconds. In other words, there is 30 seconds delay before the cell will obtain additional free resources.

Algorithm 2. Method 1 - threshold algorithm

Require: the threshold value T_{down} is set; the time W_{up} is set;

```

1: if  $N_{fc} \geq 8 + T_{down}$  then
2:   start the timer for  $W_{up}$ 
3:   while timer is on do
4:     if  $N_{fc} \geq 8 + T_{down}$  then
5:       continue the countdown
6:       if  $W_{up}$  time is up then
7:         TRX is powered off
8:       end if
9:     else
10:    stop the timer
11:   end if
12:   end while
13: end if
```

3.3 Method 2 – RRC Based Algorithm

The Method 2 is based on the Random Reference Classifier (RRC) [7] and Dynamic Ensemble Selection (DES-CD_{d-opt}) [8]. In this method, the decision to enable or disable TRX is done based on the prediction of average amount of occupied channels within the next 15 minutes. For this purpose, Multiple Classifier System (MSC) with homogeneous ensemble consisted of 20 Neural Networks (NN) (2 layers with 8 neurons each) has been created. To prevent overlearning and obtaining diversity between classifiers, each classifier was trained using randomly selected 70% of objects from the training dataset [8]. Each classifier was trained on one week data set and tested with help of another one.

The data set was prepared to fit the inputs of the NN. There are five inputs:

1. the avarage number of busy channels during last 15 minutes
2. the avarage number of busy channels during last but one 15 minutes
3. current *time* - normalized hour using Cosine function
4. current *time* - 15 minutes (also normalized)¹
5. current day - as it was shown in the Fig. 1, the traffic profiles differ on daily basis

There is only one output. The NN returns the predicted average number of busy channels for next 15 minutes. The predicted value is used as an input for the Erlang formula [9] and the number of required channels (for next 15 minutes) is calculated. On the basis of this value the number of TRXs that are needed could be obtained and the decision about turning on or off is made.

¹ each entry in the real traffic data set is corresponding to the last 15 minutes window

3.4 Research Environment

For the purpose of simulations, research environment has been prepared. The simulator was written in Java. The main task for the simulator is to generate appropriate traffic based on actual data from real telecommunication network and to manage the ongoing and incoming calls. The generator is also responsible for collecting the statistics that are needed to draw the necessary conclusions. The presentation layer and the GUI has been prepared in the form of web page. The user has the ability to run the corresponding simulations and to see the network state changes in real-time during the simulation. At the end of the simulation, appropriate set of results are presented. In the Fig. 2, the screenshots from the running simulations were shown. Each square in the Fig. 2 represents one channel in the cell. Gray color indicates that the channel is unavailable (TRX is off), red means the signal channel, green means free channel, while the yellow color represents the channel that is currently being used.

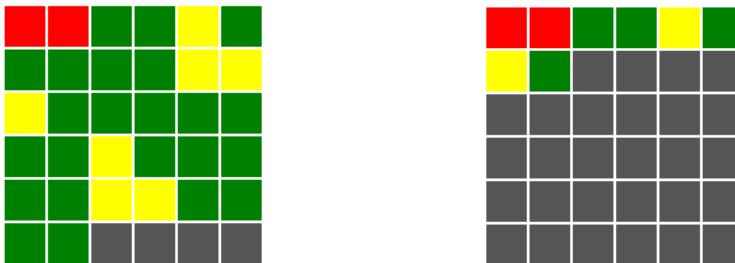


Fig. 2 Visualisation from the application. The left square represents the state of the network during high load (in the middle of the day) and the other one represents the network state at night.

4 Experiments

The experiments carried out simulations for the two methods described in the section 3. For both methods the experimental scheme was the same. Simulations were taking place for a period of 21 days. Traffic was generated by the simulator based on real data. During the simulation the following parameters were calculated:

1. *blocking* - how many calls were blocked with respect to all incoming calls
2. total number of TRXs on/off actions
3. average period between TRXs on/off actions
4. average number of active TRXs

The Method 1 tests were carried out for different values of the variables T_{down} , T_{up} and W_{up} . Thanks to this, the impact of the tested values for the final results can be noticed.

For Method 2, there is no simulation parameters.

Table 1 The results of the simulation with usage of Method 1

W_{up} [min]	T_{down}	T_{up}	<i>blocking</i>	average # of active TRXs	total # of on/off actions	average pe- riod between on/off actions
5	2	1	0.41	2.25	2384	854
	3	1	0.33	2.23	1942	1158
	4	1	0.37	2.28	1272	1384
	5	1	0.24	2.33	1153	1924
	2	2	0.22	2.38	2985	854
	3	2	0.27	2.34	2724	726
	4	2	0.23	2.39	1927	1009
	5	2	0.15	2.45	1426	1681
	2	3	0.17	2.49	4125	529
	3	3	0.16	2.42	3599	708
	4	3	0.08	2.48	2843	901
	5	3	0.05	2.51	1861	1384
	5	4	0.09	2.68	2489	719
	15	2	1	0.19	2.46	894
	15	3	1	0.29	2.53	1219
	15	4	1	0.21	2.59	691
	15	5	1	0.12	2.76	429
30	2	1	0.11	2.84	429	3892
	5	1	0.04	2.94	249	10566
60	2	1	0.05	2.78	143	8821
	5	1	0.04	2.91	157	12687

4.1 Results and Discussion

The results obtained for the simulation with Method 1 are shown in Table 1 and for Method 2 in Table 2.

Before starting to evaluate the results, the correct interpretation should be provided. Some issues should be taken into account. First, the attention should be paid to the amount and frequency of switching on/off the TRX. In fact, it is not as easy as it has been presented in this paper. Therefore, note that switching on and off the

Table 2 The results of the simulation with usage of Method 2

<i>blocking</i>	average # of active TRXs	total # of on/off actions	average pe- riod between on/off actions
0.89	2.08	149	12973

network element is relatively risky, so this action shall be made infrequently as possible, and thus in large intervals. The second parameter, which should be analyzed is the energy saving performance. In our simulation it is represented by the average number of active TRXs during the simulation. To reduce energy consumption across the network, the primary goal is to maintain an appropriate level of QoS using the minimum amount of equipment. Maintaining appropriate level of QoS means to keep *blocking* as low as possible. For the purposes of the following analysis, the 2% *blocking* is accepted and recognized as high QoS level.

In regard to the above, these results imply the following conclusions:

1. for both methods *blocking* was at an acceptable level,
2. in the Method 1 changes the value of the parameter W_{up} significantly affect the number of switching on/off actions. The difference between the minimum and the maximum value was as high as 11272 seconds,
3. in the Method 1 changes the size of thresholds T_{down} and T_{up} also significantly affect the level of *blocking*,
4. the biggest energy saving was for the Method 2 - avarage number of active TRXs was 2.18. However, there was no statistical difference between this value and the minimum value for Method 1,
5. the average time between switching on/off actions in Method 2 was significantly longer than the times obtained with Method 1 and was equal to 10438s.

5 Conclusions and Further Work

In this study the new approach for decreasing the energy consumption in mobile network was proposed. Results of the experimental investigations indicate that the proposed method with usage of RRC can be a great tool for minimizing the costs and energy consumption in ICT sector. The performed simulations also show that approach proposed in Method 2 can be used in the real world networks.

On the other hand, the most important issue for the mobile companies is to keep the QoS high. According to this, the idea of hybrid system, with multiple security levels, might be the best option. That would be also the topic for the further work for the authors. The authors would like also to implement a similar solution to improve the performance of the mobile network referring to spatial traffic changes, described in subsection 2.2.

References

1. Fettweis, G., Fettweis, G.: ICT Energy consumption - Trends and challenges. In: The 11th International Symposium on Wireless Personal Multimedia Communications, WPMC 2008 (2008)

2. Chan, C.A., Gygax, A.F., Wong, E., Leckie, C.A., Nirmalathas, A., Kilper, D.C.: Methodologies for Assessing the Use-Phase Power Consumption and Greenhouse Gas Emissions of Telecommunications Network Services. In: *The Environmental Science and Technology* 47, 485–492 (2012)
3. <http://www.datax.pl/> - company's website, which led the project [4]
4. <http://www.greennets.com> - website of the european project, which main goal was to decrease the energy consumption in ICT
5. Blume, O., Eckhardt, H., Klein, S., Kuehn, E., Wajda, W.M.: Energy Savings in Mobile Networks Based on Adaptation to Traffic Statistics. In: *the Bell Labs Technical Journal* 15, 77–94 (2010)
6. Héault, L., Strinati, E.C., Zeller, D., Blume, O., Imran, M.A., Tafazolli, R., Lundsjö, J., Jading, Y., Meyer, M.: Green Communications: A Global Environmental Challenge. In: Proc. 12th Internat. Symposium on Wireless Personal Multimedia Commun, WPMC 2009 (2009)
7. Woloszynski, T., Kurzynski, M.: A measure of competence based on randomized reference classifier for dynamic ensemble selection. In: 20th International Conference on Pattern Recognition (ICPR), vol. 1, pp. 4194–4197. IEEE Computer Press (2010)
8. Lysiak, R., Kurzynski, M., Woloszynski, T.: Optimal selection of ensemble classifiers using measures of competence and diversity of base classifiers. In: Neurocomputing (during the publication process; the paper is already accepted) (2013)
9. Freeman, R.L.: Fundamentals of Telecommunications. John Wiley (2005)

Hardware Implementation of Fourier Transform for Real Time EMG Signals Recognition

Andrzej R. Wolczowski¹, Przemysław M. Szecówka²,
Jacek Góra^{2,3}, and Marek Kurzynski¹

Abstract. Electromyography signals (EMG) are small changes of voltage which appear on the surface of a human skin as a side effect of muscles activity. These signals may be successfully used for recognition of human intention regarding the control of hand prosthesis. Such control requires quick analysis of EMG signals. One of important stages of the recognition process is the extraction of signal features. Fourier transform method is commonly used for that purpose. It assures high quality of recognition process, but its calculation is time-consuming. This study shows the construction of the experimental system of EMG signal acquisition and analysis with FFT method implemented in dedicated digital hardware. The developed architecture provides both high processing speed and small size (providing portability). Fast implementation of FFT (256-point spectrum in 3 us), competitive to signal processors, was obtained using FPGA technology.

1 Introduction

Human hand is a tool to work and struggle with, artistic emotions relay, a symbol of power. For a blind person it is also his eye and guide. Unfortunately the hand injuries are also the most common ones among the musculoskeletal-manifold injuries. In USA the amount of financial benefits granted in this respect is about 6 billion USD

Andrzej R. Wolczowski · Marek Kurzynski
Faculty of Electronics, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
e-mail: andrzej.wolczowski@pwr.wroc.pl

Przemysław M. Szecówka · Jacek Góra
Faculty of Microsystem Electronics and Photonics, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

Jacek Góra
Nokia Siemens Networks, Wrocław, Poland

annually [7]. There is strong demand for advanced multi-joint anthropomorphic device aping human hand, from both the industry interested in humanoid robots and remote hands, and the society asking for higher quality prostheses for people with disabilities. The construction issues of the artificial substitute which could reliably copy the functionality of the living original, have been extensively investigated in various research centers [1, 4, 11, 12]. Natural hand control is performed by neural impulses which stimulate its muscles. The use of these impulses to control of artificial hand is troublesome because of invasive bio/non-bio interfacing involved. For these reasons there is huge interest in other signals, that accompany natural hands movement. Muscle activity is connected with a change in electric potential between the interior of each stimulated muscle cell (myocyte) and its environment, from -60 mV to +50 mV. This change may propagate through the surrounding tissue to the surface of the skin and can be recorded there as an electromyography signal (EMG). Superposition of potential changes form a composition which depends on the geometrical configuration of the stimulated muscle cells. Therefore EMG signals carry unambiguous information on the type of muscle activity, i.e. the actual movement of the limb. Eventually the movements may be recognized through appropriate analysis of EMG signals. Muscles which control hand fingers are located in forearm and they are usually preserved after hand amputation. For these reasons EMG based control is nowadays classified among the most promising techniques, for both healthy humans operating remote hands and handicapped persons controlling their prostheses. Moreover it seems like the EMG control may be unified for these two application fields. General idea of EMG controlled cybernetic hand is presented in Fig. 1.

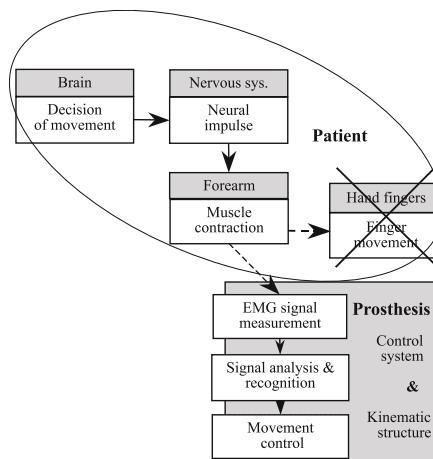


Fig. 1 Idea of EMG control of artificial hand

Taking into account various inferences accompanying the EMG signals measurements, the intention recognition is a hard problem [1,3], requiring considerable computational effort. The prosthesis providing the comfort close to the original hand

must be dexterous (adapting to the features of a grabbed object, like shape, size, elasticity, surface, etc.). This dexterity requires independent control of many degrees of freedom, what induces a huge number of recognized classes in the EMG signals, additionally increasing the already mentioned computational effort. The prosthesis should also be agile, i.e. fast enough to react accordingly to the dynamics of the environment (e.g. to grab an object which is falling down). This requires sufficiently fast control.

Eventually the contradicting requirements appear here – high complexity of calculations (to ensure dexterity), performed in a very short time (to ensure agility) and implemented in a portable, low-power device. The solution may rely on the decomposition of control algorithm to operations which may be performed concurrently, especially when implemented in the dedicated digital circuit, providing the optimal utilization of resources. A comprehensive study of state of the art in EMG control may be found in [1, 10, 13]. Analysis of scientific achievements in this domain shows that many DSP, neural and fuzzy algorithms were found feasible to recognize human intention and invoke adequate movement of an artificial hand [2]. Simultaneously these achievements are hardly found in the commercial prostheses which provide quite limited capabilities. Perhaps the reason is computational complexity mentioned above. The authors of this paper tried to solve this contradiction by development of algorithms accompanied by experiments with their implementation in dedicated digital hardware. In our previous works we have shown that algorithms like discrete Fourier transform (DFT) limited to selected points, wavelet transform (DWT) and LVQ neural networks [8] are feasible solutions which may be implemented in programmable logic devices (FPGA) in a way fulfilling mentioned requirements for speed and portability. This paper describes similar study focused on fast Fourier transform.

2 EMG Signals Acquisition and Processing

EMG signal measurements are accompanied by electrical noise that can vastly distort the information included in the signal, thus decreasing its usability for prosthesis control. The most significant noise sources are: 1) external electromagnetic fields, especially originating from devices powered by 50Hz grid; 2) high impedance of the connection between the electrode and the skin, and chemical reactions taking place there; 3) movements of the electrodes and the cables connecting them to the amplifier. The first type of noise can be effectively eliminated by means of the differential measurement circuit. Such circuit contains two signal electrodes (contacts) placed directly above the examined muscles, and a reference electrode placed above the electrically neutral tissue (preferably above the bone, joint, etc.). The signals acquired from the electrodes are subtracted one from another and then amplified. Common part of signals, containing the noise, is removed in this way, and the useful signal included in the difference is amplified. The scheme of differential measurement is presented in Fig. 1.

The remaining distortion can be essentially reduced by means of an amplifier that has a high input impedance (much higher than the skin), low output impedance, and is placed close to the signal electrodes. This is where a concept of the active electrode circuit arises – a contact electrodes which, being integrated with the preamplifier, feed the signal resistant to external noise [3]. Various electrode setups may be considered. The authors tried series of variants, with the number of electrodes varying between 4 and 12, both commercially available (Bagnoli/Delsys) [10] and in-house developed [3].

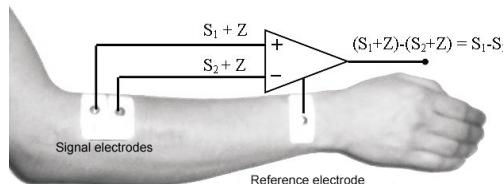


Fig. 1 Differential measurement idea; S_1-S_2 – useful signal, Z – noise

The basis for a bio-prosthesis control algorithm development is the probability density distribution of the occurrence of specified signal features for individual movement classes. This distribution can be approximated by the sets containing pairs of ÄIJEMG signal feature vector/movement class ÄI (called learning sets). During the development phase these patterns are collected from healthy humans moving their real hands. (It may be considered to use the other, healthy hand of handicapped person too). In both cases it is necessary to record the class of movement concurrently with the EMG signals. There are two basic solutions of this problem. The first one is sensoric glove which registers wrist and fingers movements. For this research alternative solution was selected - a video camera recording moving hand [9]. More sophisticated approach may be found in [6] where for the EMG signals acquisition a specially designed 8-channel EMG signals measurement circuit was applied. An overview of the whole measurement stand for identifying the relation between fingers movements and corresponding myopotentials is presented in Fig. 2. The stand includes 8 active electrodes registering the analog signals from the investigated limb, the galvanic separation block (at this point the signal is filtered and amplified), video camera for recording the moving hand image and a computer, where the signals are A/D converted, normalized and saved in a file together with video stream.

EMG signals were sampled with 1 kHz frequency and organized in 256-sample windows. This way every quarter of a second new portion of data may be analyzed to recognize currently desired action of a hand. Consequently the movement of artificial hand is a superposition of the recognized elementary actions. In this case 10 classes of movements were considered, including rotations and bends of wrist and various kinds of grasps – concentric, cylindrical, lateral. As it is shown in Fig. 3, intention recognition involves two consecutive processes: the extraction of particular signal features and, on the basis of these features, the recognition of classes

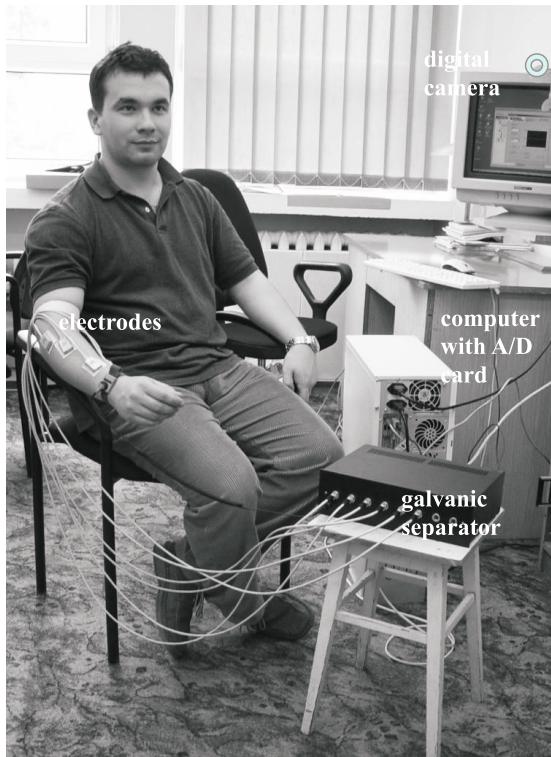


Fig. 2 Measurement stand

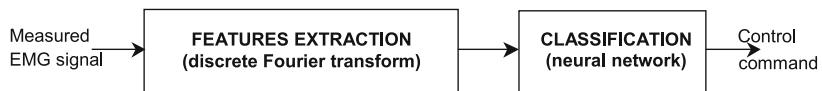


Fig. 3 Movement intention recognition block diagram

corresponding to the specific hand movements. This paper is focused on extraction of particular signal features.

Fig. 4 shows an example of EMG signal invoked by continuous movement of a hand. FFT seems quite obvious method of signal feature extraction. There are two problems however. The first one is high dimension of the features vector. 256 frequency points of spectrum multiplied by a few channels makes more than 1000 inputs for the classification algorithm, which is definitely too much, especially for neural network. Wavelet transform which delivers just a few numbers, containing comprehensive information about signal features, appears very competitive in this context. An intermediate solution is combining (adding together) groups of neighboring FFT points. Another problem with FFT is a need for efficient implementation. In this case it is estimated that after collection of sample window, which takes

256 ms, no more than 20-50 ms may be allotted for preparation of instructions for artificial hand actuators. Longer processing time would cause too large delay in prosthesis reaction making it hard to control. A fraction of this time may be used to calculate the FFT in all the channels. This requirement pushes the physical implementation platform applied for this task to the high-speed limits. There are two embedded technologies providing high performance - Digital Signal Processors (DSP) and Field Programmable Gate Arrays (FPGA). The first one is relatively easy to use. Most of DSPs may be programmed in traditional high level languages like C/C++. But its feasibility is questionable. It may be fast enough for specific algorithms, but it does not leave much space for extensions. Conceptual disadvantage is sequential manner of operation. Alternative solution is FPGA, behaving like specialized digital architecture. Its basic superiority over microprocessors is more efficient allocation of digital elements, which enables concurrent processing of the data in parallel or with pipelining. In this case the EMG channels may be handled in parallel, whilst consecutive steps of processing may form a pipeline. Disadvantage of this technology is higher design effort and more specific knowledge required for development of architecture, coding it in hardware description language like VHDL or Verilog and eventually physical implementation in FPGA. In spite of these problems, this technology was selected, due to the higher computational capabilities leaving more space for future extensions and improvements.

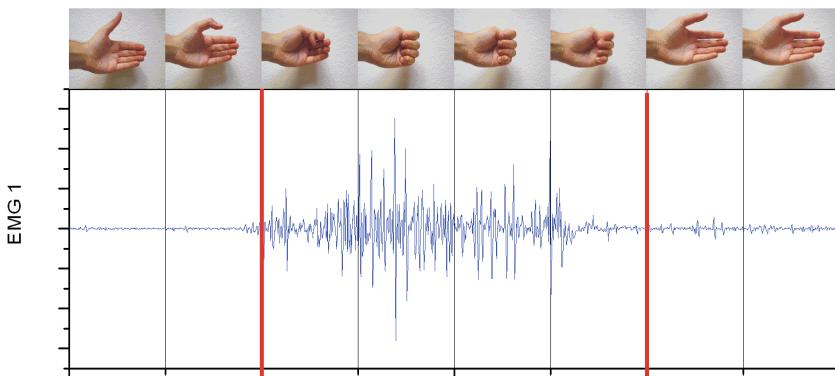


Fig. 4 EMG signals invoked by hand movement

3 Digital Hardware

Biological signals may cover a wide range of values. This feature forces digital circuitry to process both very low and very high numbers with satisfying accuracy. Fixed point data representation can hardly meet this requirement at reasonable cost. Thus the floating point representation and arithmetic shall be applied. There are several floating-point formats in use. The most popular one is defined by the IEEE 754

standard. This one provides high precision and simultaneously wide range of numbers. These advantages are achieved by very long bit representation, which leads to huge consumption of combinational logic for arithmetic operations and registers to store the numbers. The authors decided to apply their own, more scant floating point format. Bit vector consists of sign bit, 9 or 18-bit significant (always positive) and 6-bit exponent, with octal base. Table 1 shows it together with IEEE 754. It provides lower range, lower precision and much faster processing accompanied by lower consumption of FPGA resources. Series of floating-point arithmetic modules were designed by authors to process numbers in both precision formats.

Table 1 Comparison of the proposed floating-point format with the IEEE 754 standard

Parameter	Proposed format		IEEE 754	
	single precision	double precision	single precision	double precision
Full length	16 bit	25 bit	32 bit	64 bit
Significand	9+1 bit	18+1 bit	23+1 bit	52+1 bit
Exponent	6 bit	6 bit	8 bit	11 bit
Max value	5.1 E30	2.6 E33	3.8 E38	9.0 E307
Min value	1.3 E-29	1.3 E-29	1.4 E-45	5.0 E-324
Precision	2.0 E-3	4.0 E-6	1.2 E-5	2.2 E-14

Discrete Fourier transform (DFT) is a series of factors which determine the similarity of investigated signal to sine/cosine patterns of specific frequency. Fast Fourier transform (FFT) algorithm is computationally efficient version of the discrete Fourier transform. The most popular approach is based on radix-2 components (butterflies) shown in Fig. 5. In the elementary calculation two complex numbers are processed producing another two complex numbers which are then directed to the next step of computing, etc. This calculation may be performed concurrently for series of samples. Thus the hardware implementation of FFT consists of multiply and add units supported by the data flow control circuitry. The key architectural issues are sine samples delivery, optimal replication of arithmetic units and data flow organization – intermediate results must be properly delivered to blocks performing consecutive steps of computation.

An N-point radix-2 FFT consists of $N/2$ computations repeated $\log N$ times with different coefficients. It was found that full-concurrent implementation of e.g. 256-point FFT algorithm is unavailable for contemporary FPGA devices. On the other hand it is possible to create the architecture with single computational block. This solution leads to acceptable logic gates count and to lower performance as well. The most reasonable solution is to replicate the blocks to the extent allowed by FPGA capacity rather than natural complexity of algorithm. Besides arithmetic blocks, the FFT algorithm implementation requires some control logic and memory to store input and output signals, data and transform coefficients.

For the selected FPGA (1.8 million gate Xilinx Spartan 3) and 256-point FFT it is possible to apply 4 butterfly blocks together with necessary data flow control

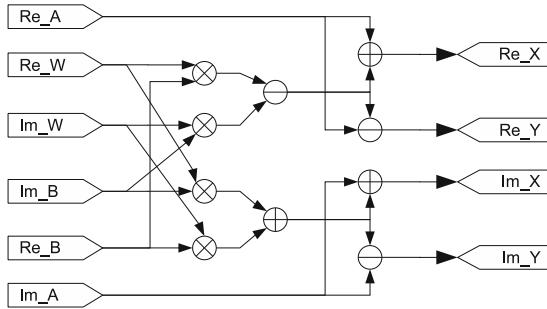


Fig. 5 Structure of the FFT radix-2 computational block (A, B – input arguments, W – transform coefficient, X, Y – results of the computation)

logic. The concept was coded in VHDL in two variants. In the first solution the chains of combinational logic were separated by registers inserted between the consecutive steps of butterfly operations. This intuitive approach led to long chains of gates between registers and consequently to low clock frequency allowed. Another approach involved more registers, inserted automatically by algorithms embedded in Xilinx tools. The result was much higher clock frequency (increasing speed) but also more clock cycles required for calculation (decreasing speed). Eventually the higher overall processing speed was achieved for the design with optimized registers - 5 times faster. Table 2. summarizes the results of synthesis and timing analysis obtained for the two variants. Both processing times – 3 and 15 us, are much better than required for operation in artificial hand (fraction of 20-50 ms). This huge margin enables application of a single FFT module for processing several channels of EMG measurements one after another. It also allows to overlap windows of samples and perform interlaced intention recognition e.g. every 128 ms instead of presumed 256 ms.

Table 2 Resource allocation and static timing analysis results for 256-point FFT implementation with 4 butterfly blocks; Spartan-3A DSP XC3SD1800A FPGA

Parameter	Intuitive design	Optimized registers (12 steps)
Registers	16 400 (49%)	19500 (58%)
LUT logic	27 200 (81%)	23 154 (69%)
DSP48A	16 (19%)	16 (19%)
Max clock freq.	18 MHz	116 MHz
Computing time	14.9 us	3.0 us

Performance of the designed circuit may be compared with other 256-point FFT implementations found in literature. Table 3 summarises computing time. Off-the shelf solutions delivered by Xilinx and Altera were used for reference in FPGA technology. For DSP, the results available for Texas Instruments and Analog Devices microprocessors are shown.

Table 3 256-point FFT computation time depending on implementation

Implementation	Authors	Xilinx design	Altera FPGA	TigerSHARC FPGA	TMS320C62x DSP	DSP
Comp. time (us)		3.03	7.56	7.70	1.55	7.90

4 Conclusions

In-house developed measurement setup for EMG signals acquisition was presented. Analysis of EMG signals is state of the art way to control artificial hand, regardless of application field – prosthesis or robotic manipulator. For prosthesis significant barrier for applying the classic DSP computations is embedded character of target device which disables use of a classic computer. It was shown however that these algorithms may be implemented in programmable logic devices (FPGA). Thus the high speed and real time processing of multichannel EMG signals is possible to achieve in portable electronics, what enables construction of next generation of hand prostheses. One of the key features of proposed solutions is specific floating point numbers bit format. Precision was decreased to the level acceptable for EMG data, bringing critical savings on both logic resources allocation and processing speed. Another crucial element was in-house development of arithmetic blocks specialized for the defined bit format. Finally the series of computational architectures was proposed. Calculation of 256-point Fourier spectrum takes 3-15 us, depending on architecture applied. The faster version outperforms off-the-shelf solutions delivered by commonly recognized vendors of FPGA devices. For multi-channel signal analysis at the highest speed, the proposed digital hardware may be replicated for each channel. For the latest FPGA circuits it would be possible to implement a few channels working in parallel in a single chip. The achieved performance may be perceived as very attractive for various metrological applications based on embedded platforms. In the context of presented research (artificial hand control) however it is not necessary. Single block may process all the channels and the huge time margin encourages to consider twice higher than presumed sampling frequency or windows overlapping. Complexity of 256-point FFT module equipped with 4 butterfly blocks was estimated to 1.2-1.4 millions of gates. For the selected FPGA providing 1.8 million gates, there is enough space for FIR filter implementation requiring 70-360 thousands of gates [5], and/or for classification algorithm, like Kohonen neural network [8]. Thus all the intention recognition tasks may be performed in a single chip. Presented results confirm that FPGA technology, in spite of relatively huge design effort involved, shall be treated as very efficient solution for smart hand prosthesis development as well as for any other metrological applications requiring high-speed multi-channel digital signal processing.

Acknowledgements. This work was financed by the National Science Center resources in 2012-2014 years as a research project No ST6/06168.

References

1. De Luca, C.J., Adam, A., Wotiz, R., Gilmore, L.D., Nawab, S.H.: Decomposition of Surface EMG Signals. *Journal of Neurophysiology* 96(3), 1646–1657 (2006)
2. Englehart, K., Hudgins, B., Parker, P.A., Stevenson, M.: Classification of the myoelectric signal using time-frequency based representations. *Med. Eng. Phys.* 21, 431–438 (1999)
3. Krysztoforski, K., Wolczowski, A.: Laboratory Test-Bed for EMG Signal Measurement, Postępy Robotyki, WKiL, Warszawa (2004) (in Polish)
4. Light, C.M., Chapell, P.H., Hudgins, B., Engelhart, K.: Intelligent multifunction myoelectric control of hand prostheses. *Journal of Medical Engineering and Technology* 26(4), 139–146 (2002)
5. Góra, J., Szecówka, P.M.: Hardware Implementation of Convolution Filters in Digital Circuits. *Elektronika* (9), 157–161 (2009)
6. Su, Y., Fisher, M.H., Wolczowski, A.R., Bell, G.D., Burn, D.J., Gao, R.X.: Towards an EMG-Controlled Prosthetic Hand Using a 3-D Electromagnetic Positioning System. *IEEE Trans. on Instrumentation and Measurement* 56(1), 178–186 (2007)
7. Syrko, M., Jablecki, J.: Quality of life-oriented evaluation of late functional results of hand replantation. *Ortop. Traumatol. Rehabil.* 12(1), 19–27 (2010)
8. Szecówka, P.M., Kowalski, M., Wolczowski, A.R., Krysztoforski, K.: Learning Vector Quantization Neural Network Implementation in a Digital Hardwar. In: Rutkowski, L. (ed.) Computational Intelligence: Methods and Applications, pp. 81–91. Exit, Warsaw (2008)
9. Wolczowski, A.R., Myslinski, S.: Identyfing the relation between finger motion and EMG signals for bioprosthetic control. In: Proc. 12th IEEE Int. Conf. on Methods and Models in Automation and Robotics, Miedzyzdroje, pp. 817–822 (2006)
10. Wolczowski, A.R., Kurzynski, M.: Human-machine interface in bioprosthetic control using EMG signal classification. *Expert Systems* (27), 53–70 (2010)
11. RSL Stepper, The hand (2013), <http://www.bebionic.com>
12. Ottobock, Users and Patients. Arm Prosthetics (2013),
<http://www.ottobock.com>
13. Zecca, M., Micera, S., Carrozza, M.C., Dario, P.: Control of Multifunctional Prosthetic Hands by Processing the Electromyographic Signal. *Critical Reviews in Biomedical Engineering* 30(46), 459–485 (2002)

Data Preprocessing with GPU for DBSCAN Algorithm

Piotr Cal and Michał Woźniak

Abstract. One of the well known density-based clustering algorithm is DBSCAN, which is commonly used for class identification in spatial databases. In this paper we propose its modification which could shorten the global computational time by introducing a simple data preprocessing with graphics processing unit (GPU). The GPU consists of many small cores which allow parallel computation using single instruction multiple data model (SIMD). Combining it with CPU power of computing system can improve application performance. In order to estimate the computation gain of our proposition compared with the original one we carried out a set of experiments on an artificial dataset.

Keywords: clustering, unsupervised learning, parallel computation, GPU computing.

1 Introduction

Clustering is the part of unsupervised machine learning. Its task is to organize data into groups (clusters) using a given similarity measure. Clustering can give us information about data structure. Because computer technology develops fast and currently the multicore systems are easy to obtain, we can use them for improving algorithms to get better performance.

There are many methods to cluster data and some survey and categorization of algorithms were done [1], [10]. Very popular density-based algorithm is a DBSCAN

Piotr Cal · Michał Woźniak
Wrocław University of Technology,
Department of Systems and Computer Networks,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: {piotr.cal,michal.wozniak}@pwr.wroc.pl

(density-based spatial clustering of applications with noise) [5], [9]. The advantages of DBSCAN is that it does not need knowledge about number of clusters and is very resistant to noise. The main idea for DBSCAN is to check neighborhood (with given ϵ) for each objects. If it contains at least the specified number of objects the cluster is expanded (cluster is expanded if density is bigger than a given threshold). The time complexity of this algorithm is $O(n^2)$ if input data do not use any indexing structure. However, as many other clustering algorithms, it needs to compute similarity measure for each of objects. If the data dimensionality is high, an algorithm can run too slow.

The attempts to improve DBSCAN by parallel implementation are in a work [11]. Authors use a master-slave technique and 3 steps procedure. First data set is divided and obtained partitions are sent among processors. Then all processors locally perform DBSCAN algorithm and eventually the results are merged. This solution uses R*-tree structure for quick data access. Many other master-slave methods are in works [3], [7], [6]. In a paper [8] authors show the parallel solution using disjoint-set data structure. This allows them to perform merge operation concurrently.

The reason why we try to gain better performance comes from the real life. The amount of information grows up rapidly and to handle it we need more computing power. Nowadays the power of single processing unit does not increase as fast as before because of physical barriers like speed, size or heat output. The cost of doubling the performance of the one processor is now unprofitable comparing to the cost of using two processors. However, the development of computers deals with it using multicore systems. One of the example of such a computing component is GPU. Original the GPU was designed to process graphics but now it can be an usefull tool for parallel computing. It consists of hundred or thousand small cores which can be used for general purpose computation (GPGPU).

GPU has been used for clustering task by other researches. In [4] the k-means algorithm using GPU was introduced. Also DBSCAN with GPU finds interesting solution in [2]. The main steps for parallel clustering are core searching and expand cluster procedures. Additionally, authors propose parallel similarity-join and indexing-structure operation for the fast similarity search. They show a very good speedup on the large artificial datasets.

Both the GPU and CPU can work together to solve the clustering problem. The main objective of this paper is to verify whether this hybrid architecture may speedup the DBSCAN algorithm. The evaluation of our method is made by a comparison to the sequential version of algorithm. The next section shortly describe DBSCAN algorithm and the GPU modification for data preprocessing which is the subject of this work. Then we mention the implementation and we show experiment results. The last section describes conclusions and future work.

2 Algorithm

2.1 DBSCAN

DBSCAN is a density based clustering algorithm. As input user defines two parameters: ϵ and minimum point number (*minPoint*). On the basis of them algorithm computes the distance between two objects and if its value does not exceed ϵ , the objects become neighbors. The algorithm defines three basic objects types. The core is an object that has at least the given *minPoint* neighbors. The boarder is an object that is not a core and has at least one neighbor that is a core. Other objects are called noises.

The algorithm starts looking for an unlabeled core point and after that it begins the procedure of creating cluster. It often uses a stack technique. Initial object gets a new label for the whole cluster. All its neighbors are labeled with the same identifier and they are pushed to the stack if they were unlabeled. Next the algorithm takes object from the stack and repeats the process. After the stack is empty, the one cluster is build and again, the algorithm is looking for a new unlabeled core point. It is finished if there are no more unlabeled core points.

2.2 GPU Data Preprocessing

If we do not have any indexing structure with data, the algorithm has to visit one object once and for this object it checks distance between all other objects. The most time consuming procedure is searching for neighbors. However, information about neighbors and the types of objects can be obtained in parallel by graphic processing unit (GPU). After that, the host processor has got crucial information for the fast sequential computing. The Fig 1 presents a difference in computation between sequential method (CPU) and method with parallel data preprocessing using GPU.

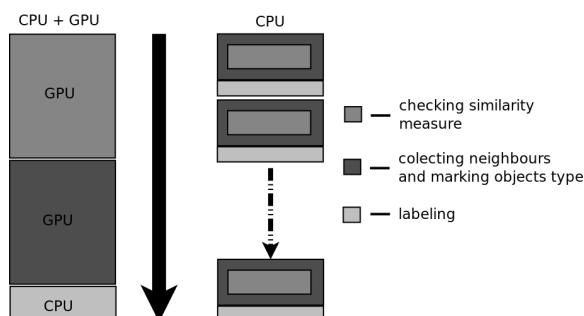


Fig. 1 The difference in computation for sequential and GPU supported DBSCAN algorithm

Algorithm 1 Neighborhood matrix creation

Require: n – number of objects; A – neighborhood matrix (size n^2) initialize with 0
 id – processor id; k – processors number
DO PARALLEL (id , A)

```

while  $id < n^2$  do
     $src = Ent(\frac{id}{n})$ ;  $dest = id \bmod n$ 
    if  $src < dest$  then
         $sym = dest \cdot n + src$ 
         $\hat{A}[id] = DISTANCE(src, dest)$  // return 1 if value is less than  $\epsilon$ 
         $A[sym] = A[id]$ 
    end if
     $id = id + k$ 
end while
```

The first action in our application is sending data to the GPU memory. The process of sending information needs time but it is relatively short comparing to the rest of computation. After data is sent, GPU starts computing in parallel the distance between objects and, if it is lower than ϵ , it puts positive value into the neighborhood matrix. The procedure is presented in listing (2). Because this matrix is symmetric, it only needs to perform this action $\frac{n^2}{2}$ times. After neighborhood matrix is created, application checks all rows in parallel collecting and counting neighbors to assign the type of object. This procedure is presented in listing (1). At this state the neighborhood matrix is modified and it keeps a list of neighbors for all objects. Next, we send results to host where sequential DBSCAN is performed using knowledge about neighbors and objects types.

Algorithm 2 Neighbor collecting

Require: n – number of objects; A – neighborhood matrix (size n^2)
 id – processor id; k – processors number
DO PARALLEL (id , A)

```

while  $id < n$  do
     $offset = id \cdot n$ ;  $sum = 0$ 
    for  $i \in 0 : n - 1$  do
        if  $A[i] > 1$  then
             $A[sum] = i$ 
             $sum = sum + 1$ 
        end if
    end for
     $A[sum] = -1$  //denotes the end of neighbors
     $id = id + k$ 
end while
```

2.3 Implementation

The application was created using C/C++ language and CUDA/C API for GPU programing. C is the low level programming language and it lets reach a good performance. Also C language is supported by CUDA architecture. Both procedures,

neighborhood matrix creation and assigning the type of object are done in parallel using GPU device. However GPU architecture is difficult and for optimal performance it needs some code optimization. It is important to exploit blocks and threads to maximize cache memory usage. In our implementation we do not use advanced method so the application still could be improved.

Another important thing is that we do not focus on the memory usage. Memory is also important to process big datasets. However, we use the same memory usage for both sequential and GPU version for comparison. The lack of memory for large data is caused by neighborhood matrix which needs n^2 fields.

3 Experiment

The main goal of the experiment is to verify the possible speedup of proposed method and their evaluation by comparison to the sequential version of algorithm.

3.1 Set-Up

We use artificial generated dataset which contains two dimensional continuous feature. In this case, the value of similarity measure is Euclidean distance between two objects. All tests were carried out on two machines. First, NVIDIA GeForce GT 645M graphics cards (denotes as GeForce) and Intel core i5 CPU with 64-bit operating system (CPU i5 (64)) and second with NVIDIA Tesla C2050/2070 (Tesla) and Intel core i7 CPU with 32-bit operating system (CPU i7 (32)). The time was measured always for proposed and sequential method for the same dataset. The first test shows dependency between DBSCAN computation time and data set size. The next one verifies if the parameters of algorithm have any impact on computation time.

3.2 Results

Using GPU for data preprocessing does not change algorithm output. The solution was exactly the same for sequential method and GPU one. The reason is simple. We use parallel computation only to prepare data for quicker sequential computation so the algorithm methodology is the same in both cases. The difference is that we take off a lot of computation from CPU and pass them to GPU for parallel processing. Fig 1 shows samples of output after DBSCAN algorithm.

The Fig 2 presents the times of computation for sequential (CPU) method and method supported with GPU. Times for all method rise exponentially. It is caused by the complexity of algorithm which is n^2 . The best result was achieved by Tesla GPU and the worst by CPU i7 (32). Tesla machines are designed for better computation performance and GeForce are designed for better graphics processing. Also Tesla

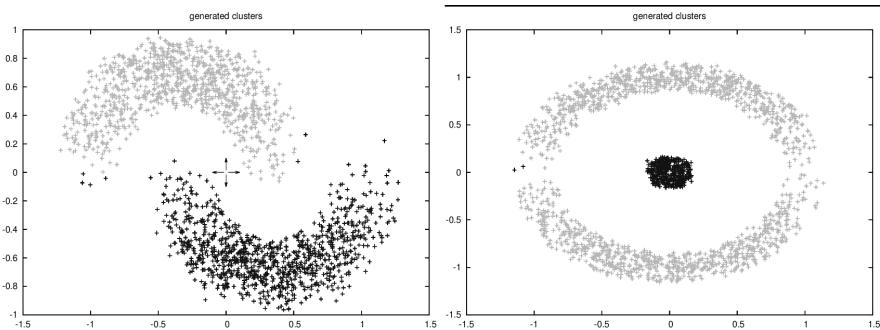


Fig. 1 Example of generated two dimensional data after clustering process. Results are the same for sequential and GPU supported method.

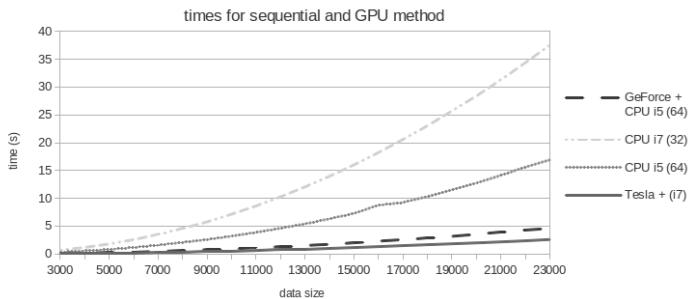


Fig. 2 Times of computation

has more single cores than Geforce GPU. This caused that Tesla is almost twice faster. Times for CPU's are several times higher. The CPU i5 with 64-bit operating system has better results than CPU i7 with 32-bit operating systems. The reason of such a situation is the architecture of operating system. The architecture 64-bit operating system processes arithmetics operation faster than 32-bit. We should mention that after data preprocessing the sequential part of computation (labeling) was done in insignificant time comparing to the total time of algorithm. Due to this, we focus on parallel data preprocessing. For larger data or data with some structures the distribution of computation could be different.

The Fig 3 shows the speedup depending on the data size. This value is more or less fixed and data size does not affect it. Again, the reason of this is that the algorithm is the same but the part of computation is done in parallel on GPU with more cores. Comparing to other mentioned works, this value is not as big as it could. This is caused by simple implementation. The GeForce is 3.5 times faster than CPU i5 (64) and Tesla is about 14 times faster than CPU i7 (32). However, Fig 3 shows that we can easily estimate times for larger datasets.

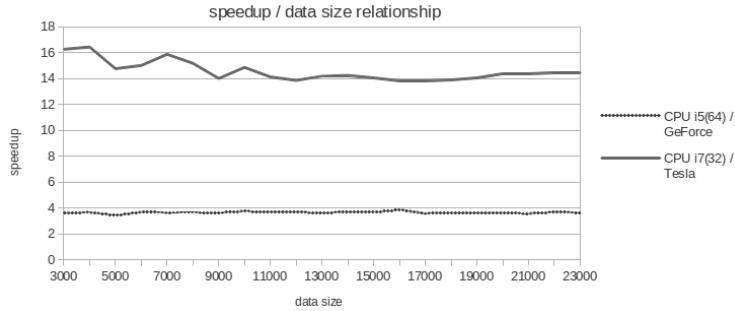


Fig. 3 Speedup / dataset size relationship

The Fig 4 presents how time depends on parameter ϵ value. When ϵ is bigger, time slightly increases. This is caused by the higher number of neighbours, which causes an increase in computation, especially for host when it creates clusters. The second parameter *minPoint* did not cause any noticeable changes in time. The parameters are important to build good clusters but for computation time they do not make essential difference.

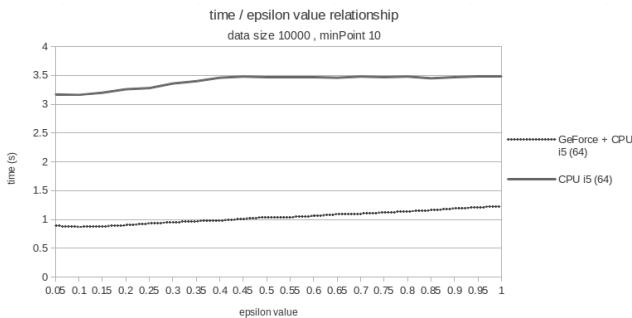


Fig. 4 Time / epsilon relationship

4 Conclusions

The application shows that for the artificial data sets we can see speedup using GPU to prepare data for DBSCAN algorithm, what is not surprising. The value of speedup depends on used graphic cards and for more powerful GPU it could be better. Our solution, simple data preprocessing in parallel, improve the performance for data without any structure. The experiment shows that the most needed computation power part is the process of looking for neighbors. It is strongly connected

with commuting similarity between objects. If objects have more dimensions, the parallel computation will be even more effective. The application at the beginning prepare data on GPU. The data given to host consist of neighbors list and objects type. Results show that for sequential part it is sufficient for high speed clusters creation. Our solution is not complicated and better results with GPU were achieved in work [2]. However we are planning to combine their idea with our application to improve the results.

In the future we are going to optimize the GPU implementation for better performance and we want to adjust the application to make it more memory efficient, what will allow to carry out experiments on very big data sets. We are planning to combine solutions from our application and the other work [2]. Also we want to add idea from the work [8] where authors decomposed data using disjoint-set data structure for multicore system. It should let us perform clustering algorithm in parallel from the beginning to the end and it significantly increases performance using distributed system with multiple CPU and GPU devices.

Acknowledgements. This work is supported by the Polish National Science Center under a grant N N519 576638 for the period 2010-2013.

References

1. Andreopoulos, B., An, A., Wang, X., Schroeder, M.: A roadmap of clustering algorithms: finding a match for a biomedical application. *Briefings in Bioinformatics* (2009)
2. Böhm, C., Noll, R., Plant, C., Wackersreuther, B., Zherdin, A.: Data Mining Using Graphics Processing Units. In: Hameurlain, A., Küng, J., Wagner, R. (eds.) *Transactions on Large-Scale Data- and Knowledge-Centered Systems I*. LNCS, vol. 5740, pp. 63–90. Springer, Heidelberg (2009)
3. Brecheisen, S., Kriegel, H.-P., Pfeifle, M.: Parallel density-based clustering of complex objects. In: Ng, W.-K., Kitsuregawa, M., Li, J., Chang, K. (eds.) *PAKDD 2006*. LNCS (LNAI), vol. 3918, pp. 179–188. Springer, Heidelberg (2006)
4. Cao, F., Tung, A.K.H., Zhou, A.: Scalable clustering using graphics processors. In: Yu, J.X., Kitsuregawa, M., Leong, H.-V. (eds.) *WAIM 2006*. LNCS, vol. 4016, pp. 372–384. Springer, Heidelberg (2006)
5. Ester, M., Kriegel, H.P., Jörg, S., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise, pp. 226–231. AAAI Press (1996)
6. He, Y., Tan, H., Luo, W., Mao, H., Ma, D., Feng, S., Fan, J.: Mr-dbscan: An efficient parallel density-based clustering algorithm using mapreduce. In: Proceedings of the 2011 IEEE 17th International Conference on Parallel and Distributed Systems, ICPADS 2011, pp. 473–480. IEEE Computer Society, Washington, DC (2011)
7. Li, H., Chen, M., Gao, X.: Parallel dbscan with priority r-tree. In: *Information Management and Engineering, ICIME* (2010)

8. Patwary, M.A., Palsetia, D., Agrawal, A.: A new scalable parallel dbscan algorithm using the disjoint-set data structure. In: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, SC 2012, pp. 62:1–62:11. IEEE Computer Society Press, Los Alamitos (2012)
9. Sander, J., Ester, M., Kriegel, H.-P., Xu, X.: Density-based clustering in spatial databases: The algorithm gdbSCAN and its applications. In: Data Min. Knowl. Discov., pp. 169–194 (1998)
10. Xu, R., Wunsch, D.: Survey of clustering algorithms. IEEE Trans. Neural Netw. 16(3), 645–678 (2005)
11. Xu, X., Jäger, J., Kriegel, H.-P.: A fast parallel clustering algorithm for large spatial databases. In: Data Min. Knowl. Discov., pp. 263–290 (1999)

Part VII

**Pattern Recognition and Image Processing
in Robotics**

Structured Light Techniques for 3D Surface Reconstruction in Robotic Tasks

M. Rodrigues, M. Kormann, C. Schuhler, and P. Tomek

Abstract. Robotic tasks such as navigation and path planning can be greatly enhanced by a vision system capable of providing depth perception from fast and accurate 3D surface reconstruction. Focused on robotic welding tasks we present a comparative analysis of a novel mathematical formulation for 3D surface reconstruction and discuss image processing requirements for reliable detection of patterns in the image. Models are presented for a parallel and angled configurations of light source and image sensor. It is shown that the parallel arrangement requires 35% fewer arithmetic operations to compute a point cloud in 3D being thus more appropriate for real-time applications. Experiments show that the technique is appropriate to scan a variety of surfaces and, in particular, the intended metallic parts for robotic welding tasks.

1 Introduction

One the main challenges in automated robotic tasks is the development of flexible systems that can be set up quickly and easily with minimum user intervention and can be switched over to another product line while maintaining quality and productivity. Small and medium enterprises normally find it difficult to invest in new

M. Rodrigues · M. Kormann
GMPR – Geometric Modelling and Pattern Recognition Group,
Sheffield Hallam University, UK
e-mail: {m.rodrigues, m.kormann}@shu.ac.uk

C. Schuhler
TWI – The Welding Institute, Cambridge, UK
e-mail: clement.schuhler@twi.co.uk

P. Tomek
MFKK – Invention and Research Center Services, Budapest, Hungary
e-mail: peter.tomek@mfkk.hu

technologies requiring expert knowledge and extensive human training. Addressing those issues, the MARWIN Project [8] (Decision Making and Augmented Reality Support for Automatic Welding Installations) funded by the EU offers a solution to human-robot interaction by developing a cognitive welding robot where welding tasks and parameters are intuitively selected by the end-user directly from a library of CAD models. Robot trajectories are then automatically calculated from the CAD models and validated through 3D scanning of the welding scene. The role of the user is limited to high level specification of the welding task and to the verification of welding parameters and sequences as suggested by MARWIN.

This paper focuses on describing optical techniques for fast 3D reconstruction including image processing methods and a comparative analysis of sensor configurations concerning mathematical formulation and computational requirements. Optical methods for 3D surface reconstruction can be divided into two categories: passive methods that only require images from fairly uncontrolled illumination, and active methods that require controlled patterns of light to be projected onto the scene [10]. Passive methods include stereo vision, Shape-from-X (SfX) and Simultaneous Localization and Mapping (SLAM). Active methods include Time-of-Flight (ToF) and Structured Light where surface patches are reconstructed from the geometric relationships existing between the source of light and sensor devices. Our research is focused on structured light techniques using coded and uncoded light patterns (e.g. [2, 15]). The main advantages of our techniques are speed and accuracy as a surface is scanned and reconstructed from a single 2D image in 40ms.

A review of structured light techniques is provided in [11] and a classification of projection patterns is proposed in [16]. A number of patterns exist such as the Kinect Box based on projecting dot patterns [17], uncoded stripes [13], coded stripes [15], colour patterns [4], and alternating stripe patterns or phase shifting [18]. Concerning physical arrangement, structured light scanners assume an angled configuration in which the normals from the centre of the projector and the centre of the camera sensor meet at the calibration plane with a number of methods being proposed for reliable reconstruction (e.g. [1, 3, 5, 7, 12]). In this paper we further our previous work on fast 3D reconstruction using structured light [14] by comparing a novel configuration and corresponding mathematical formulation where the light source and camera sensor are parallel to each other against a standard angled configuration.

2 The Geometry of Structured Light

Structured light scanners project patterns of light onto the target surface whose image is recorded by a camera. The detected pattern in the image is combined with the spatial relationship between the light source and the camera, to determine the 3D position of the surface. Many coding schemes have been proposed, a review is given in [16] and a few examples are depicted in Figure 1. In order to avoid taking multiple shots as in time-multiplexing, colour or variable width stripes can be used. In this research we use grey-level stripes as in uncoded stripes of Figure 1 in which

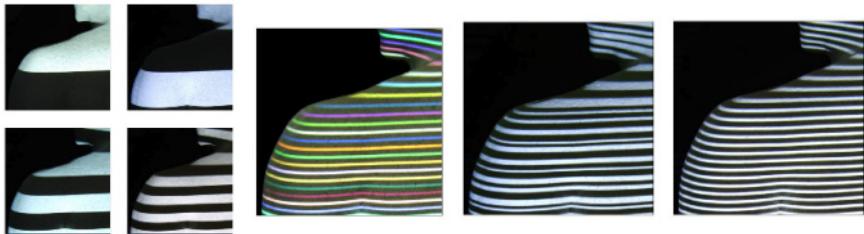


Fig. 1 Examples of coding schemes. From left to right: time-multiplexing, colour coding, variable width, and uncoded stripes.

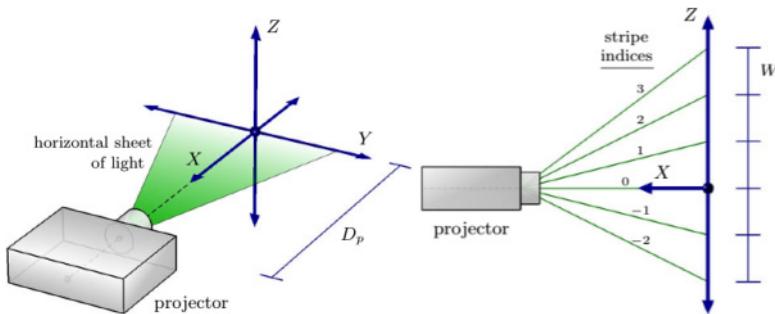


Fig. 2 A coordinate system is defined in relation to the light source

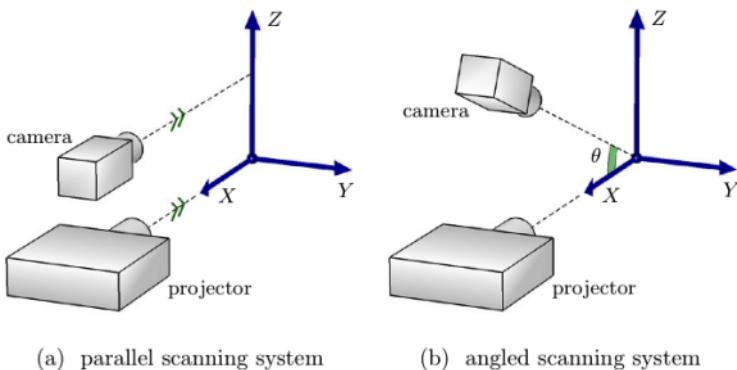


Fig. 3 Left, parallel configuration; right, standard angled configuration

every stripe has the same width allowing a maximum resolution with a sufficiently dense pattern of stripes.

In order to proceed to the mathematical formulation, first it is necessary to define a coordinate system. We choose it to be in relation to the light source as shown in Figure 2: the origin is set at a known distance from the centre of the projection.

Concerning the physical arrangement, the camera and projector can be either in parallel or angled configuration as illustrated in Figure 3. The problem we are trying to solve is defined as follows. Every pixel in the image as captured by the camera needs to be mapped into 3D to the chosen world coordinate system. This is solved by determining to which light plane or stripe each pixel belongs to, and then through trigonometric relationships find the coordinates of the surface point imaged by the pixel.

Assuming that image processing functions are available to detect stripes and their respective indices in the image (see Section 3), the mapping from the image sensor to a surface point is defined by the pixel location and its corresponding stripe index. From the point of view of the image sensor as illustrated in Figure 4, the mapping of a pixel to its coordinates \mathbf{p} is defined as $\mathbf{p} = \mathbf{c} + (0, -hPF, vPF)$ [14]. Here \mathbf{c} is the centre of the image, F is the focal length of the camera, P is the pixel size and each pixel in the sensor plane is a square of size $PF \times PF$.

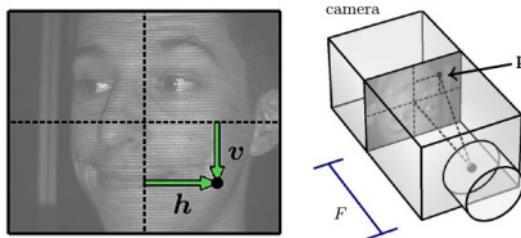


Fig. 4 The position of a pixel lying on a stripe of index n in the image plane is transformed to coordinates (h, v) then to a point \mathbf{p} .

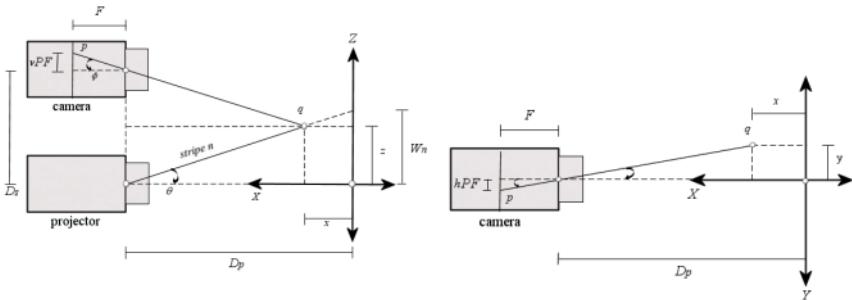


Fig. 5 Mapping of a surface point q from (h, v, n) in camera space to (x, y, z) in system space. Left, side view of scanner; right: view from the top.

The mathematical formulation for the parallel configuration as depicted by the geometry of Figure 5 is derived entirely from trigonometric relationships between the various measured distances [14]. The following parameters are defined:

D_s :	the distance between the camera and the projector
D_p :	the distance between the projector and the system origin
W :	the width between successive stripes in the calibration plane
P :	the pixel size in the sensor plane of the camera
F :	the camera focal distance
v, h :	vertical and horizontal position of the pixel in camera space
x, y, z :	3D coordinates of a surface point
n :	stripe index from the light source
θ, ϕ :	angle between a light plane and its location in camera space

The expressions to compute the coordinates (x, y, z) in world or system space from a pixel location (v, h) on stripe n (mapping to a point on the surface of the scanned object) is defined by the geometry of Figure 5 as:

$$x = D_p - \frac{D_p D_s}{v P D_p + W_n} \quad (1)$$

$$y = \frac{h P D_p D_s}{v P D_p + W_n} \quad (2)$$

$$z = \frac{W_n D_s}{v P D_p + W_n} \quad (3)$$

The mathematical formulation for the angled configuration has originally been described in [13]. In this case, the mapping of a pixel (v, h) on stripe n to its corresponding surface point (x, y, z) can be written as:

$$x = \frac{D_c v P + W_n (\cos \theta - v P \sin \theta)}{v P \cos \theta + \sin \theta + \frac{W_n}{D_p} (\cos \theta - v P \sin \theta)} \quad (4)$$

$$y = h P (D_c - x \cos \theta - z \sin \theta) \quad (5)$$

$$z = W_n (1 - \frac{x}{D_p}) \quad (6)$$

where the expressions for x in equation (4) can be substituted into equation (6) for z , and then into equation (5) for y , so that each is formulated as a function of v , h and n .

The resolution in 3D space (in millimetres) for both parallel and angled configurations are mathematically equivalent: it depends on the spacing between the projected planes, on the distance between the surface and the light source and the dimension of the pixel space. The vertical resolution is affected by the spacing between the projected stripes which occupies several pixels in the image plane. The horizontal resolution measured along each stripe is dependent on the horizontal dimension of the pixel space. We can formalize the resolution (δ) along the (x, y, z) dimensions by [14]:

$$\delta x = \frac{PD_p^2 D_s}{[vPD_p + W_n][(v+1)PD_p + Wn]} \quad (7)$$

$$\delta y = P(D_p - x) \quad (8)$$

$$\delta z = \frac{W}{D_p}(D_p - x) \quad (9)$$

In practice, it is possible to process one vertex in 3D per image pixel along each stripe (in the horizontal direction assuming that stripes are horizontal). To illustrate the magnitude of resolution, for an IDS camera with 1280×1024 pixels calibrated at 300mm from the object's surface, the measured pixel dimension at the calibration plane is 240 microns. Across stripes, the peak-and-through distance of a stripe is 5 pixels. Therefore this means that along stripes the horizontal resolution is 240 microns (0.24mm) and across stripes the vertical resolution is $240 \times 5 = 1200$ microns (1.2mm).

From equations (1)–(6) the parallel configuration requires less computing cycles than those of standard angled scanners: 24 arithmetic operations per vertex in 3D against 37 operations for angled (assuming that $\cos \theta$ and $\sin \theta$ are pre-computed). Thus, the parallel configuration design is more appropriate for 3D real-time processing, as it requires 35% fewer operations. In terms of mechanical construction and calibration, both designs are equivalent; the significant difference is how the surface points (x, y, z) are calculated as described above from the known calibrated (i.e. measured) values of extrinsic parameters D_s , D_p , W and θ .

3 Image Processing

In order to reconstruct a surface in 3D according to the mathematical definitions in the previous section, the following steps are necessary: (1) image noise filtering; (2) detect stripe patterns in the image; (3) index the stripes in relation to the centre of the system; (4) map to 3D space. The mapping to 3D is accomplished using equations (1), (2), (3) for a parallel configuration or equations (4), (5), (6) for an angled configuration.

Figure 1 shows a raw image with detail on the right (raw and filtered). Image noise removal operations are required normally median and mean filters. Details of raw and filtered image are shown on the right – each stripe on the bottom right has a smooth Gaussian profile making it easier to detect the peak of each stripe with sub-pixel accuracy, which is determined as the local maxima in the greyscale intensity values of each stripe. Once peaks or local maxima are detected, algorithms are run (e.g. [2, 13] based on maximum spanning tree and flood filling) to detect each stripe correctly indexing them, where the stripe with index 0 (zero) corresponds to the light plane emanating from the centre of the projector (the reference projected light plane). Stripes with ascending positive indices are above the centre stripe, negative

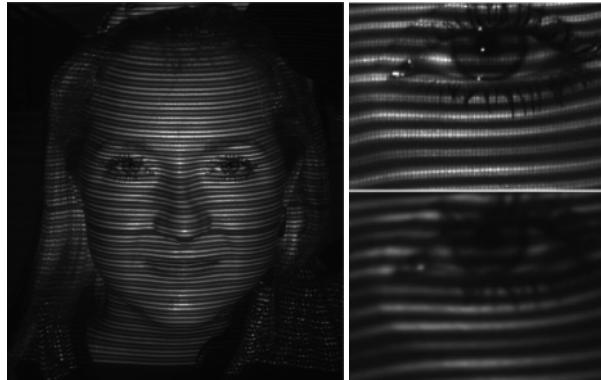


Fig. 1 Left, raw image; top right, raw detail; bottom right, filtered detail.

below. With the correct indices determined, the mapping given by equations (1)–(3) or (4)–(6) are then applied resulting in a point cloud of vertices in 3D space.

4 Experimental Results

We developed a flexible design of a compact scanner prototype with internal adjustable mechanisms allowing it to be configured and calibrated either as a parallel or angled configurations. To this end, we must choose a calibration plane at an arbitrary distance from the light source (200–800mm for the design) and adjust the camera until the line perpendicular to the image sensor is either parallel to the light source or until it intercepts the system origin at the calibration plane as shown in Figure 3(b) for an angled configuration. Note that the calibration plane does not set the maximum distance an object can be scanned. The limitation is due to the brightness of the projector and the reflective properties of the surface being scanned which, ultimately, will determine whether or not stripe patterns can be detected.

The prototype depicted in Figure 1 uses an IDS UI-1241LE camera board [6] with image size 1280×1024 and a Microvision PicoP laser projector [9] with resolution 848×480 pixels. Tests were carried out on the actual metallic parts used in the robotic assembly of the MARWIN project together with wood and skin surfaces. The difference between the parallel and angled configurations is the number of arithmetic operations to evaluate a point cloud; the calibration procedures are equivalent, and the 2D image processing functions are the same and depend on the surface being scanned. Equally, if 3D post-processing is required such as hole filling and smoothing, these functions are also equivalent for both configurations.

Figure 2 depicts several reconstructed surfaces. As 3D reconstruction maps a 2D pixel to its counterpart 3D vertex, texture mapping is a straightforward process and there is no need for a post-processing texture mapping step as in standard scanners. Due to high specularity, metallic parts require two distinct image filtering functions

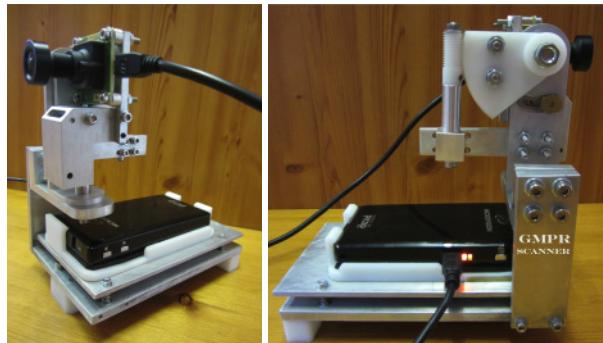


Fig. 1 The compact GMPR Scanner assembly with laser projector and CMOS camera. Dimension $W \times H \times L = 100 \times 150 \times 150$ mm.

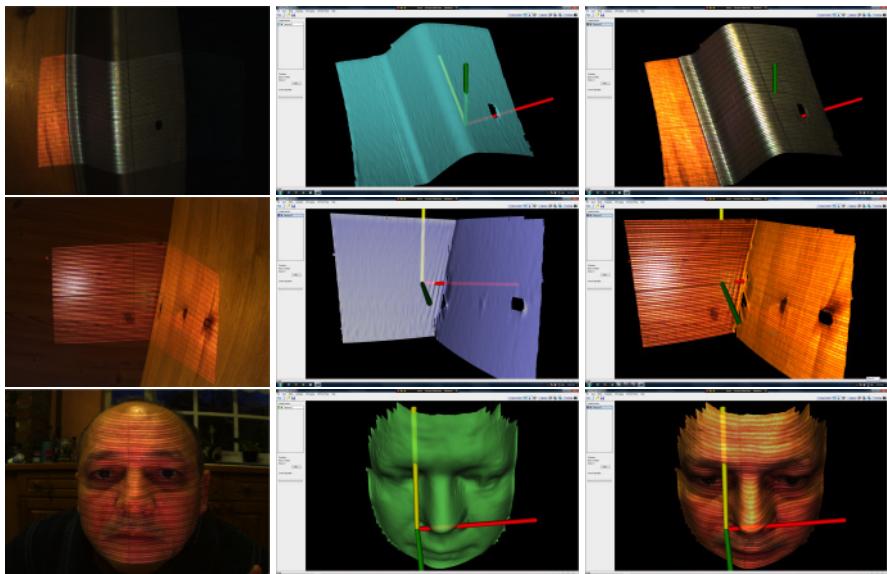


Fig. 2 From top to bottom: metal, wood, skin. Left column: 2D image; middle: 3D reconstructed surfaces; right: surface with texture mapping.

namely a median followed by a weighted mean filter to make stripes detectable. In contrast, the other surfaces only require either a weighted mean or a mean filter. In terms of overall processing time we can break down in several steps: 40ms for 2D image filtering (median and mean filters), 40ms for calculation of a point cloud in space for the angled configuration (26ms for the parallel configuration), and 120ms for 3D mesh pos-processing (triangulation, small hole filling, mesh repairing and

smoothing). With all the times computed, the current prototype in either parallel or angled configuration can operate in real-time at about 5 frames per second.

5 Conclusions

This paper describes research as part of the MARWIN project on 3D scanning of metallic components for robotic welding tasks. The focus is on the development of structured light scanners; a novel mathematical formulation for a parallel configuration is presented and compared to a standard angled arrangement of light source and image sensor. The built prototype consists of a camera board and a laser projector mounted on a mechanism designed for fine adjustments yielding either parallel or angled configurations. All interfaces to the control computer, camera boards and projector use standard USB connectors.

The main image processing requirement is identified as filtering images such that pixel intensities across stripes describe a smooth Gaussian profile for reliable detection of stripes peaks and throughs. Our experience has demonstrated that for reflective surfaces such as metals, employing a median filter followed by weighted mean filters can effectively remove extremes of noise resulting in the required Gaussian profile of pixel intensities. On less reflective surfaces such as wood and skin tissue, a combination of mean and weighted mean filters proved effective.

The significance of the design shown here is that it provides a simple to build and calibrate scanner that can be used for a variety of tasks and surfaces. In particular, its small form factor and interfaces easily integrate within robotic tasks, such as welding and assembly of metallic components of the MARWIN project. Future work includes registration and fusion of scanned welding components and tests with LED light sources in the visible and near-infrared spectra and a detailed error analysis of scanned surfaces.

We acknowledge EU funding to the MARWIN Project (Decision Making and Augmented Reality Support for Automatic Welding Installations), contract no. 286284, *Research for the Benefit of SMEs*, from Nov 2011 to Oct 2013. We also acknowledge Willie Brink for his work on geometry.

References

1. Albitar, C., Graebling, P., Doignon, C.: Robust structured light coding for 3D reconstruction. In: Int. Conf. Comp. Vis (ICCV), pp. 1–6 (2007)
2. Brink, W., Robinson, A., Rodrigues, M.A.: Indexing Uncoded Stripe Patterns in Structured Light Systems by Maximum Spanning Trees. In: Proc. BMVC British Machine Vision Conference (2008) ISBN 978-1-901725-36-0ss
3. Chen, S., Li, Y.: Vision processing for realtime 3D data acquisition based on coded structured light. IEEE T. Image Process. 17, 167–176 (2008)

4. Clancy, N.T., Stoyanov, D., Groch, A., Maier-Hein, L., Yang, G.Z., Elson, D.S.: Spectrally-encoded fibre-based structured lighting probe for intraoperative 3D imaging. *Biomedical Optics Express* 2, 3119–3128 (2011)
5. Gorthi, S., Rastogi, P.: Fringe projection techniques: Whither we are? *Opt. Laser Eng.* 2, 133–140 (2010)
6. IDS Imaging GmbH, USB2 uEye LE Camera Boards (2013),
www.ids-imaging.com (Last accessed January 9, 2013)
7. Kawasaki, H., Furukawa, R., Sagawa, R., Yasushi, Y.: Dynamic scene shape reconstruction using a single structured light pattern. In: *IEEE Int. Conf. Comp. Vis. and Pat. Rec. (CVPR)*, pp. 1–8 (2008)
8. MARWIN: Decision making and augmented reality support for automatic welding installations, EU-funded Project no. 286284, Research for the Benefit of SMEs, from November 2011 to October 2013, <http://www.marwin-welding.eu/>
9. Microvision Inc., MicroVision SHOWWX+ Laser Pico Projector (2013),
microvision.com (Last accessed on January 9, 2013)
10. Mirota, D.J., Ishii, M., Hager, G.D.: Vision-based navigation in image-guided interventions. *Annu. Rev. Biomed. Eng.* 13, 297–319 (2011)
11. Mountney, P., Stoyanov, D., Yang, G.Z.: Three-dimensional tissue deformation recovery and tracking. *IEEE Signal Proc. Mag.* 27, 14–24 (2010)
12. Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsiontas, V., Chamzas, C.: Methods for 3D digitization of cultural heritage. *Journal of Cultural Heritage* 8, 93–98 (2007)
13. Robinson, A., Alboul, L., Rodrigues, M.: Methods for Indexing Stripes in Uncoded Structured Light Scanning Systems. *Journal of WSCG* 12(1-3), ISSN 1213–6972
14. Rodrigues, M., Kormann, M., Schuhler, C., Tomek, P.: An Intelligent Real Time 3D Vision System for Robotic Welding Tasks. In: *IEEE ISMA 2013 The 9th Int. Sym. on Mechatronics and Apps.*, Amman, Jordan, April 9-13 (2013)
15. Rodrigues, M., Robinson, A.: Novel methods for real-time 3D facial recognition. In: Sarrafzadeh, M., Petratos, P. (eds.) *Strategic Advantage of Computing Information Systems in Enterprise Management*, pp. 169–180. ATINER, Athens (2010)
16. Salvi, J., Pagès, J., Batlle, J.: Pattern codification strategies in structured light systems. *Pattern Recognition* 37(4), 827–849 (2004)
17. Smisek, J.: 3D with Kinect. In: *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, November 6-13, pp. 1154–1160 (2011)
18. Wang, Y., Liu, K., Lau, D.L., Hao, Q., Hassebrook, L.G.: Maximum SNR Pattern Strategy for Phase Shifting Methods in Structured Light Illumination. *J. Opt. Soc. Am. A* 27(9), 1962–1971 (2010)

Multi-modal People Detection from Aerial Video

Helen Flynn and Stephen Cameron

Abstract. Automatic people detection is a highly desirable capability for autonomous Unmanned Aerial Vehicles (UAV) operating in search and rescue missions. Here we present an approach for automatic people detection from aerial video using a combination of infrared and visible imagery. The results presented show that existing state of the art part-based people detection algorithms can be used with considerable success in aerial imagery. In a cluttered urban environment, the detection rate is approximately 80% over an entire flight pattern. We show that, by fusing visible and infrared imagery and tracking detections over time, the false positive rate is reduced to a minimum.

1 Introduction

There has been great interest in the use of small robotic helicopter vehicles over the last few years. Although there are regulatory issues involved in flying these that are still to be solved, they have the potential to provide a practical mobile aerial platform for a small fraction of the cost of a conventional manned helicopter. One potential class of applications for these are in searching for people. Police helicopters are regularly used for this purpose in the UK and elsewhere, where a trained human ‘spotter’ will typically use a combination of infra-red (IR) and visual light imaging to seek people lost or hiding from the authorities. The equipment used for this purpose on manned helicopters is typically large, heavy and expensive, and relies on the spotter to detect the targets. Our purpose here is to explore the use of a new generation of cameras that may be suitable for this task on small unmanned aerial vehicles (UAVs). Furthermore, communications between such a UAV and a human controller may be unreliable or limited, and so we would also like to be able to use

Helen Flynn · Stephen Cameron
Dept of Computer Science, Univ. of Oxford
e-mail: {helen.flynn, stephen.cameron}@cs.ox.ac.uk

new computer hardware and image processing techniques to at least partially automate the detection process; we would much prefer to have a UAV send back just images of a potential target to the controller for verification, rather than expecting the controller to continually stare at a screen for long periods.

The detection of people from small UAVs poses several challenges, including

- **Payload.** Small research UAVs can typically carry less than a kilogram of load, depending on their endurance (flight-time). This has to account for cameras, lenses, mounts, and computing and communication equipment. Although a mission-ready UAV might manage more, hard decisions will have to be made regarding the use of payload. In particular, heavy zoom lenses are unlikely to be viable.
- **Image shake.** Small vehicles are particularly susceptible to wind gusts, and there is normally little weight allowance available to provide complete dampening of mechanical vibrations within the UAV.
- **Processing power.** Although techniques such as GPU programming may be very effective for image processing, the overall computing power available on a UAV will be limited.
- **Infra-red picture quality.** ‘Traditional’ quality IR cameras are heavy, particularly if they use active cooling systems. We are using instead a micro-bolometer camera, which weighs just 80g including a single lens (but not including all of the required processing electronics). Such cameras are solid state, and the price may be expected to fall; but the most expensive (c. \$10K) model currently available outside of the military provides an image resolution of only 640×480 .
- **Weather conditions.** To be fit for purpose a UAV-based solution should be able to deal with at least some variety of weather.

Moreover, current state of the art people detection algorithms have not yet been tested in detecting people from above. One of the aims of this work was to see if they would be applicable. Given that these algorithms are prone to false positives especially in cluttered scenes, we employed infrared imagery to narrow down the search and to increase the confidence in detection.

The work presented here forms part of a project¹ whose focus lies in the creation and control of swarms of autonomous helicopter UAVs that are deployed in such a scenario to sense the environment in the most efficient way possible and to report their findings to a base station on the ground. These UAVs would typically be flying at a height of approximately 10-20m.

2 Related Work

There has been considerable interest in visual people detection, but primarily using fixed camera locations. For these the ‘targets’ are generally upright — so in a standard orientation — and there are good algorithms available that utilise basic vision

¹ SUAAVE, <http://www.suaave.org/>, EPSRC project EP/F064179/1

algorithms such as Haar features with AdaBoost [11] or the Histograms of Oriented Gradients (HOG) detector [3].

A handful of work to date has focussed on the problem of people detection from aerial imagery. Oreifej et al. [9], for example, employ the use of HOG detectors in fairly simplistic scenarios where the people are fully visible and in standard upright poses. The approach presented in [10] combines thermal and optical imagery, exploiting also the geometric constraints of the shadows cast by humans in order to increase the detection accuracy. Gaszczak et al. [6] also use thermal imagery but from a fixed wing aircraft at a relatively high altitude, using Haar classifiers on thermal signatures and Gaussian shape matching as secondary confirmation. Andriluka et al. [1] appear to be the only ones to have used more sophisticated state of the art people detectors on aerial imagery. They use the part-based detectors of [2] and [4] in order to detect people in more challenging poses using images acquired from a quad rotor UAV flying in an indoor office environment.

3 Method

Visual light cameras provide high resolution over three colour channels, whereas the raw output of an IR camera is monochromatic (apparent surface temperature) and generally of poorer resolution. However high resolution visual light images can take considerable time to process; most image processing algorithms take time that is at least linear in the image size. We therefore decided to consider an approach in which the IR images are used to find interesting regions in the image, namely bright (warm) regions of an appropriate size. Such regions can be found quickly, and each are then used to centre a search window for a vision processing algorithm to scan for evidence of a person.

In the work reported in this paper we already have some video camera footage from such a vehicle; this was originally used to assess the feasibility of spotting people using visual image processing only. UAVs are a new technology and experimental UAVs can still lose control and fall from the sky. For this reason — and for convenience — we have based the results in this paper from images taken from suitable fixed vantage points instead. However the equipment has been sourced with a view to flying it on a UAV in early 2013, to confirm the results described here.

3.1 Overview

We acquired video footage from a handheld camera rig comprising a visible camera (Point Grey Chameleon 1280×960 with a 8.5mm lens and 2/3" sensor, giving field-of-views of 55°×42°) and an infrared camera (Thermoteknix Miricle Microcam 640×480 with a 18.8mm lens and 1/3" sensor, giving field-of-views of 46°×35°). Footage was recorded in various cluttered urban scenes looking down to the ground from a height of approximately 20m with the cameras pointing roughly 45° to the

horizontal. The video streams were sampled at approximately 15 frames per second and stored in a laptop computer for later processing.

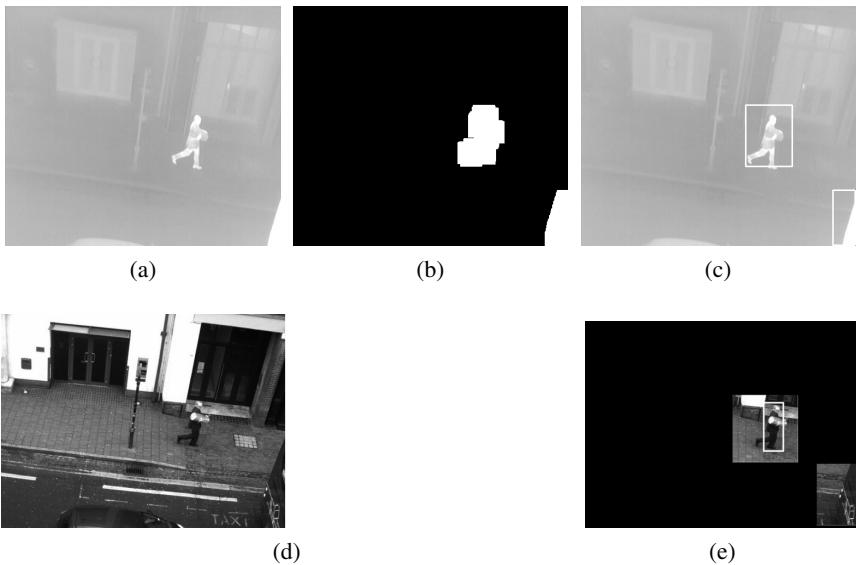


Fig. 1 (a) and (d) show the original infrared and visible images. (b) shows the result of thresholding the infrared image and dilating the resulting hot regions found. (c) shows the bounding boxes of the contours found in (b). (e) shows the result of running a person detector over the corresponding regions in the visible image.

The two cameras were aligned spatially such that it was possible to compute a reasonably accurate homography between the two images, which maps one image plane to the other. The infrared image is used to narrow down the search to parts of the scene more likely to contain humans. Since humans exhibit heat more than the environment, they can be expected to appear as brighter regions in the thermal image. We therefore threshold the thermal image to identify the hot spots and then find the corresponding regions in the optical image. These are analysed using a people detection algorithm as discussed in Section 3.3. Figure 1 shows the effect of the image processing pipeline.

3.2 *Infrared Processing*

Once the appropriate IR regions are identified it is necessary to find the corresponding region in the optical image.

3.2.1 Homography Estimation

A homography is a projective mapping which maps one image plane to the other. The relationship between points in both images is given by

$$\mathbf{p}_{optical} = H\mathbf{p}_{thermal} \quad (1)$$

where H is a 3×3 matrix and p is an image point in homogeneous coordinates $[x, y, 1]^\top$. We choose two images both rich in detail so that there are many corresponding points. The homography is computed using a least squares method which minimises the sum of squared reproduction errors after projecting points in one image into the other [7]. For our experiments, we computed the homography in advance by manually selecting a set of corresponding points in both images, and passing this to a least squares minimisation solver to estimate H .

3.2.2 Blob Extraction

The histogram of an infrared image is typically Gaussian-shaped, representing the ambient temperature of the environment. Humans tend to be hotter than their surroundings, so their intensity should appear to the right of and far away from the main (contrast normalised) distribution. We therefore threshold the image at a point near the end of the intensity range, say > 250 (out of 255), leaving us with a set of white blobs that represent the hottest parts of the scene. The infrared signature of a human tends to produce a set of disjoint hotspots, representing the exposed skin regions, since the body is not a uniform temperature if they are wearing heavy clothing. We therefore perform a morphological dilation on the hotspots which gives us larger hotspots as shown in Figure 1b. Projecting the bounding box of each blob into the optical image using homography H gives a set of higher confidence regions of interest within which to search.

3.3 People Detection

There is already a huge body of research on the detection of people within images, though little work has been done on the detection of people in aerial imagery. These algorithms can broadly be classified into monolithic and part-based approaches.

Monolithic models use a single descriptor to encode the appearance of an object. Detection is performed by scanning an image at multiple positions and scales, with each window passed to a discriminative classifier which makes a decision about the presence or absence of an object. One of the most popular and successful models for people detection is the Histogram of Oriented Gradients (HOG) detector. In this algorithm, histograms of image gradients are computed and normalised in local overlapping blocks which are then concatenated to form a single descriptor. A linear support vector machine is used to classify the detection window. The HOG detector

tends to learn the overall shape of a person, and works well when the person is fully visible and in a similar pose to what the classifier was trained on. It does not work when the person is highly articulated or partially occluded.

Part-based models give the flexibility required to deal with highly varying body poses. Most of the state of the art part-based detectors are built on the pictorial structures framework [5]. Here, an object is represented as a flexible configuration of parts, where one such configuration is denoted by $L = \{l_0, \dots, l_N\}$, with l_i denoting the location of part i . The posterior over part configurations L given image evidence E is calculated using Bayes' rule: $p(L|E) \propto p(L)p(E|L)$. This framework assumes that the overall likelihood can be decomposed into the product of individual part likelihoods. Under this assumption, the configuration posterior factorises as:

$$p(L|E) \propto p(l_0) \cdot \prod_{i=0}^N p(E|l_i) \cdot \prod_{(i,j) \in G} p(l_i|l_j) \quad (2)$$

Sum-product belief propagation is applied in order to compute the marginal posterior of the torso $p(l_0|E)$, which is then used to determine the bounding box.

In this work we use the discriminatively trained part-based model of Felzenszwalb et al [4]. In this model, configuration of body parts is modelled as a star-shaped model, which has a root part (torso) and several other parts (typically limbs and head) each of which is connected to the root part. Training of the model is semi-supervised, in that the locations of the parts are not manually annotated as with other models, but are instead inferred from the data using visually salient reoccurring structures in the training data. The configuration of body parts which maximises Equation 2 is found with max-product belief propagation. The entire model is trained in a discriminative fashion aiming to maximise the margin between positive and negative examples. The appearance of each part is modelled using support vector machines with HOG descriptors, while a deformation cost penalising the part for deviating from its ideal location relative to the root is obtained using gradient descent. Since this algorithm doesn't need manual annotation of parts, it can be trained on a much larger training set.

One of the main aims of this work was to evaluate the applicability of various state of the art people detectors on aerial images. Such imagery poses more of a challenge than the more traditional settings in which these algorithms have been tested. We tested the standard HOG detector and the part-based detector of Felzenszwalb; it should be noted that the Felzenszwalb detector runs much more slowly than the HOG detector for the same image size.

3.4 Tracking

Minor differences between frames can mean that the detector may correctly detect a person in one frame and fail to detect it in the next, and there can be spurious false alarms. Consequently, per frame detection performance tends to be hit and miss.

We therefore evaluate the performance of our method over an entire flight pattern. By tracking the detections over time, we can be more confident of the detection hypothesis. Consistent detections raise confidence while random detections can be ignored, and this reduces the overall false positive rate.

We compute the homography between each pair of consecutive frames. Given a location in the previous frame, the homography is used to compute the corresponding location in the current frame. Each detection in the previous frame is projected into the current frame, and the bounding boxes of the current and previous detections are compared. If a previous detection overlaps significantly with a current detection, this is taken to be a repeated detection, and this increases the confidence in that detection. Only those detections with a high confidence score are retained.

4 Results

Detections within an urban environment are shown in Figure 1. These are the results of applying Felzenszwalb’s part-based detector to the visible image; we found that the HOG detector was not competitive in this setting due to the wide variation in poses. Unsurprisingly, using the infrared image to narrow down the search space reduces the false positive rate dramatically. Per frame detection tends to be sporadic, but by tracking high confidence detections over time, the overall detection rate is over 80% for the four scenarios in our experiments. At this detection rate the false positive rate is about 20%.

Weather and the type of environment affects the success of detection. Figure 2a presents an interesting example of an image taken on a hot day where the roof in the foreground shows up as a hot region in the infrared image. Using our method, the corresponding region in the visible image is searched, and this can result in false positives. This can be counteracted by only considering isolated infrared blobs of the right size. The downside of this is that in very hot weather, the ground around a person is likely to be very hot also, so that their infrared signature is not clearly segmented.

Using the tracking method allows us to impose a higher threshold on an acceptable score for the detection, because if a person not detected in every single frame, it will be detected in at least some of them. This helps in reducing false positives. Currently we are only retaining information from the previous 5 frames.

Homography computation between successive pairs of 1280×960 frames takes on average 1.64s on a Intel quad core 2.83GHz CPU with an integrated graphics card. Processing the infrared image and scanning the reduced image space with the Felzenszwalb detector takes approximately 2.4s. This compares very favourably with the use of the same detector over the entire image, which takes around 15s per frame, and it suggests that the algorithm should run in ‘real-time’ (say, under 100ms per frame) when optimised and run on the hardware expected to be available before long.

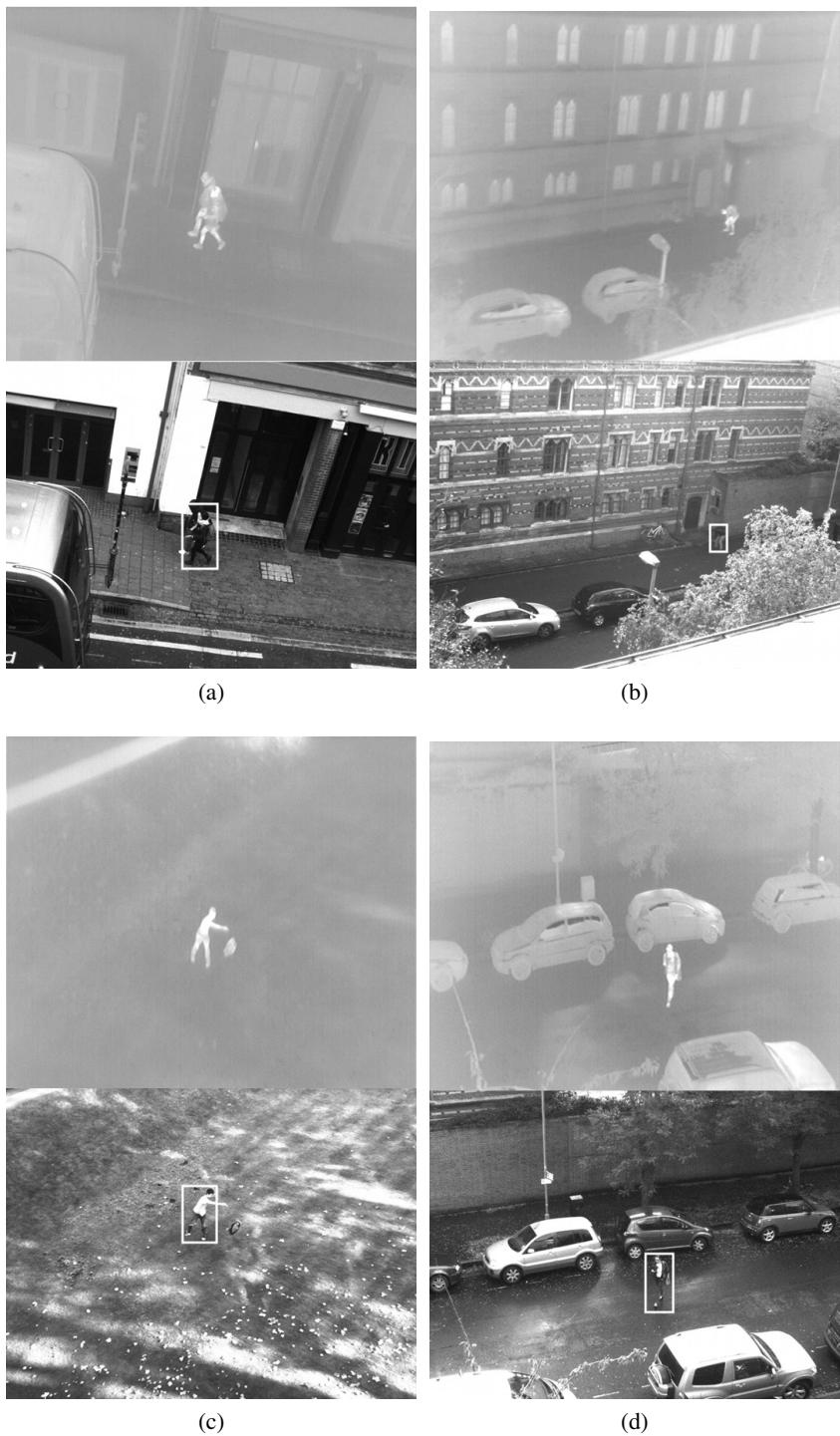


Fig. 1 Detections within an urban environment



Fig. 2 Some false positives

5 Conclusions

A method of people detection from aerial video has been proposed which combines infrared and visible imagery to increase the accuracy of detection and reduce false positives. It has been shown that state of the art people detection algorithms can be applied to aerial imagery with considerable success. The methodology has been tested in difficult environments with people in non-standard poses and has been shown to be promising as we have detected around 80% of people with minimal false positives. Furthermore, the detection times suggest that this method can be employed in real-time on the sort of hardware expected to be available in the near future. This is work-in-progress, and we are continuing to work on refining the method, and in preparing to test it running on a real UAV. Future work in this area will include: only scanning images at the most likely scales in which a person can appear; doing some image processing on the infrared image itself; and estimating the relative velocity of the person and using that to predict where to search in the next image, in order to increase confidence over multiple frames.

References

1. Andriluka, M., Schnitzspan, P., Meyer, J., Kohlbrecher, S., Petersen, K., von Stryk, O., Roth, S., Schiele, B.: Vision Based Victim Detection from Unmanned Aerial Vehicles. In: International Conference on Intelligent Robots and Systems, Taipei, Taiwan (2010)
2. Andriluka, M., Roth, S., Schiele, B.: Pictorial Structures Revisited: People Detection and Articulated Pose Estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
3. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: Proc. Conf. Computer Vision and Pattern Recognition, pp. 886–893 (2005)
4. Felzenszwalb, P.F., McAllester, D., Ramanan, D.: A Discriminatively Trained, Multi-scale, Deformable Part Model. In: IEEE Conference on Computer Vision and Pattern Recognition (2008)
5. Fischler, M.A., Elschlager, R.A.: The Representation and Matching of Pictorial Structures. *IEEE Transactions on Computer* 22(1), 67–92 (1973)
6. Gaszczak, A., Breckon, T.P., Han, J.W.: Real-time People and Vehicle Detection from UAV Imagery. In: Proc. SPIE Conference Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques, vol. 7878, No. 78780B (2011)
7. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2004) ISBN: 0521540518
8. Huang, C., Wu, B., Nevatia, R.: Robust Object Tracking by Hierarchical Association of Detection Responses. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part II*. LNCS, vol. 5303, pp. 788–801. Springer, Heidelberg (2008)
9. Oreifej, O., Mehran, R., Shah, M.: Human Identity Recognition in Aerial Images. In: Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, pp. 709–716 (2010)
10. Rudol, P., Doherty, P.: Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery. In: Proc. IEEE Aerospace Conf., pp. 1–8 (2008)
11. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, pp. 511–518 (2001)

The Classification of the Terrain by a Hexapod Robot*

Adam Schmidt and Krzysztof Walas

Abstract. This paper presents a new approach to the terrain classification by a hexapod robot using the tactile information. The data was acquired using the force/torque sensor mounted on the walking robot foot. Two types of classifiers were used and compared: the Normal Bayes Classifier (NBC) and the Classification And Regression Tree (CART). The article comprises the description of the experimental setup followed by the presentation of feature selection process and the comparison of the two classifiers' accuracy. The classification system presented in the article allows the walking robot to recognize the type of the terrain on which it is currently walking on with over 90% accuracy.

1 Introduction

The ability to identify the terrain is crucial for the mobile robots to work efficiently outdoors. The information about the type of the terrain allows the robot to adapt the motion control algorithms to the changing traction conditions. The terrain identification can be obtained by using visual data as it was presented in [4, 13]. However, relying only on the visual cues may be misleading. For example robot might be looking at the wallpaper or patterned artificial flooring. Tactile perception is required in order to obtain a reliable terrain recognition. The majority of the research concerning this topic focuses on the wheeled mobile robots which are using the vehicle chassis acceleration measurements to characterize the terrain. Such approaches were

Adam Schmidt · Krzysztof Walas

Poznan University of Technology, Institute of Control and Information Engineering
e-mail: {adam.schmidt, krzysztof.walas}@put.poznan.pl

* K. Walas is funded by the Polish National Science Centre, Grant No. 2011/01/N/ST7/02070.

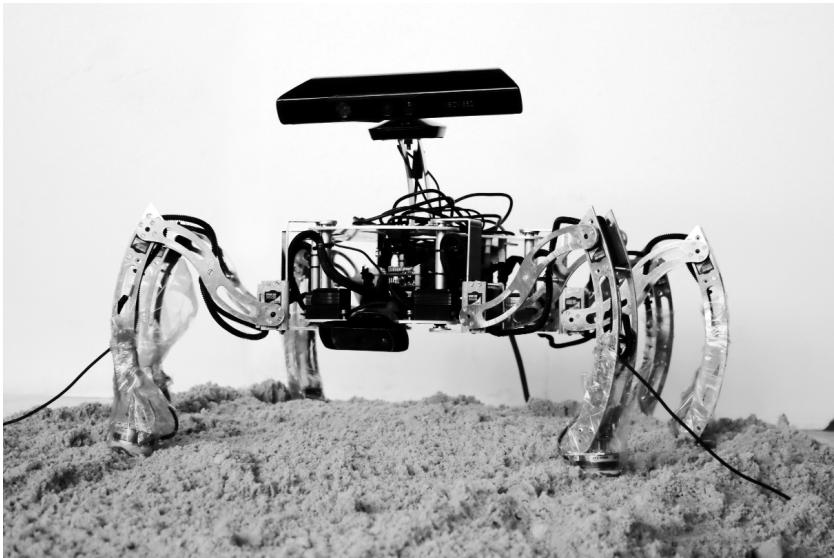


Fig. 1 Messor robot while walking on the sand

described in [3, 12, 15]. Different approach was presented in a series of articles [6–8] where the authors used simple tactile probe for terrain classification.

Furthermore, the information about the terrain is even more indispensable for legged robots. The first attempts to obtain the tactile information about the terrain were described in [10, 11]. Some later advances were presented for the RHex, robot with rotating legs, in [5]. The most recent research on this topic was presented in [9] where the authors used a single vibrating leg of the robot to classify the terrain. In our work we would like to go beyond the state of the art and obtain an on-line (after one step) tactile classification of the ground type.

The paper is structured as follows. First, the legged robot and experimental setup is described. Afterwards, the feature selection process and the classification of the terrain based on the force/torque signals evaluation are presented. Finally, the concluding remarks are given followed by future work plans.

2 Experimental Setup

2.1 Robot

The six-legged robot walking robot Messor was used in the experiment. It is equipped with ATI Force/Torque Mini-45 sensor mounted on its front foot. The robot is shown in Figure 1. Its trunk is 30.6 cm long and 26 cm wide. The segments of the leg are: coxa 5.5 cm, femur 16 cm, tibia 23 cm. The robot is described in details in [14].

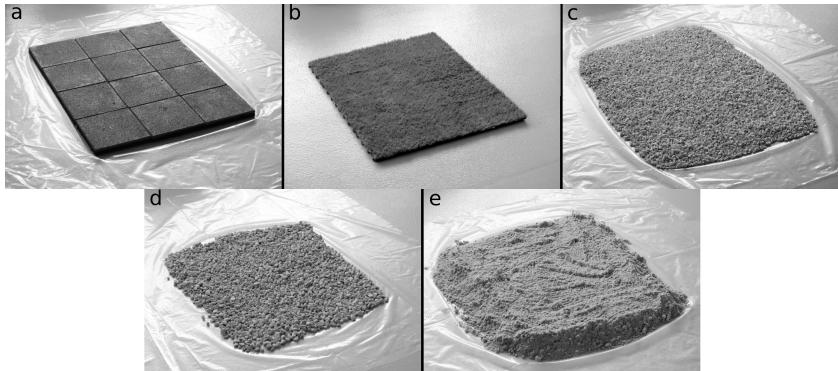


Fig. 2 The types of terrain used in experiments: soft ground (a), artificial grass (b), gravel (c), pebbles (d), sand (e). Solid ground – concrete floor surrounding the terrain probes.

2.2 *Terrain Types*

Six different terrain types were considered in the study. The ground types were presented in Figure 2. The experimental set consists of: soft ground – rubber paver tiles, artificial grass, solid ground – concrete floor, gravel, pebbles and sand. The solid ground is the concrete floor surrounding the terrain probes. The selected materials are characterized by different damping, friction and bounciness.

2.3 *Data Registration*

The data was registered for each terrain type separately. The robot made 10 trials of 6 steps for each ground type. Two different walking speeds were used. The data was registered with the sampling frequency of $f_s = 200\text{Hz}$ and consists of 6 signals, 3 force vector components (F_x , F_y , F_z) and 3 torque vector components (T_x , T_y , T_z). The registered sequences were manually divided into steps. Each step was truncated to the length of 400 samples for the first walking speed $V_1 = 2.95 \text{ cm/s}$ or 900 samples for the second walking speed $V_2 = 1.32 \text{ cm/s}$. The registered data for the slower movement with the walking speed V_2 is shown in Figure 3. Each subplot contains a different force/torque signal and each series of the subplot represents the response registered for a different terrain type.

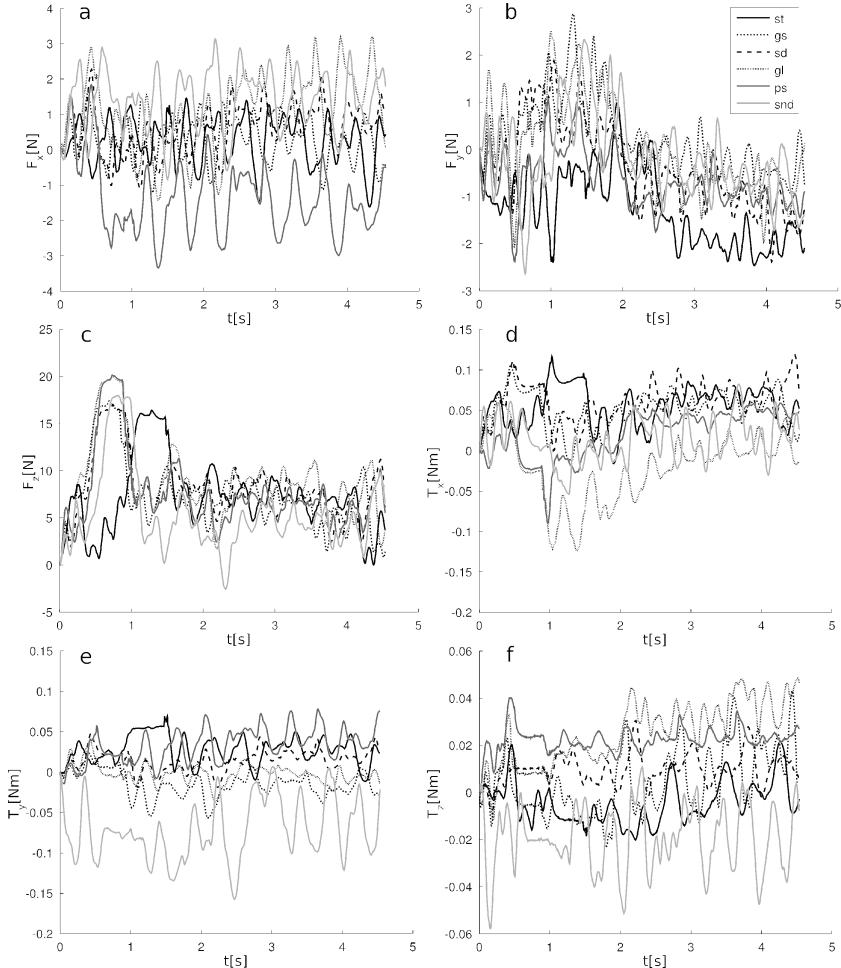


Fig. 3 The force (a,b,c) and torque (d,e,f) signals of the 6-DOF generalized contact force vector registered for walking speed $V_2 = 1.32 \text{ cm/s}$. Single step for each type of terrain: solid line – soft ground (st), dotted line – artificial grass (gs), dashed line – solid ground (sd), short dash line – gravel (gl), gray line – pebbles (ps) and light gray – sand (snd).

3 Classification and Results

3.1 Feature Vector

Each step is described by 6 signals which consist of either 400 or 900 samples. As the dimensionality of such data is too high to use it directly as an input of the classification process a reduced features vector was proposed. Similarly to the approach

presented in [8, 15] the feature vector consisted of statistical parameters of each of the recorded signals:

$$x = [\mu_1 \mu_1^2 \dots \mu_1^6 A_1 \text{Kurt}_1 \mu_2 \dots \text{Kurt}_6] \quad (1)$$

where μ_i is the mean value of the i -th channel, μ_i^j stands for the j -th order central moment of the i -th channel while A_i and Kurt_i are the skewness and kurtosis of the i -th channel. Thus, the complete feature vector consists of 42 values.

Three samples sets were created. The first called $D4$ consists of feature vectors extracted from the 400-samples signals. Analogously, the $D9$ was created using the 900-samples data. The $D49$ consists of both the previous sets.

3.2 Variable Importance

In order to evaluate the importance of particular variables in the feature vector a full, not pruned CART [1, 2] trees with surrogate variables were constructed for all the samples sets. The information gained by variables during the creation of the tree (including the surrogate variables) can be used to rank the variables according to their importance. The variables used in the creation of the tree along with their relative score as well as the mean of the importance scores are presented in the table 1.

It is worth noting, that both the mean and the variance of each channel were the discriminating features. The feature importance scores were used to define feature subvectors to be used during the classification evaluation. The $x_{D4}^{0\%}$ consists of features whose importance was bigger than 0% in the $D4$ set, analogously subvectors $x_{D4}^{5\%}, x_{D4}^{10\%}, x_{D9}^{0\%}, x_{D9}^{5\%}, x_{D9}^{10\%}, x_{D49}^{0\%}, x_{D49}^{5\%}$ and $x_{D49}^{10\%}$ were defined. Additionally, subvectors $x_\mu, x_V, x_{\mu+V}$ consisting of the mean values, variances and both the means and variances of all the six channels were created.

3.2.1 Normal Bayes Classifier

The leave-one-out approach was used to evaluate the classification accuracy of the Normal Bayes Classifier [1]. The cross-validation procedure was performed for each possible combination of the feature vector and the samples set. The classification accuracy is presented in the table 2. It is clearly visible that the best results were obtained when the feature vector consisted of the mean values of all the channels.

3.2.2 CART

A similar experiment was performed using the CART as a classifier. The classification accuracy is presented in the table 3. The CART classifier in general performed worse than the NBC. However, it is worth noting, that satisfactory results were obtained using the combination of means and variances as the feature vector.

Table 1 Variable importance

<i>D4</i>	<i>D9</i>	<i>D49</i>		mean	
Variable Importance					
μ_2^3	18%	μ_6^2	20%	μ_5	17%
μ_2	14%	μ_3	17%	μ_6	11%
μ_3^3	11%	μ_3^4	11%	μ_3	9%
μ_5	11%	μ_1^2	9%	μ_4	8%
μ_5^5	9%	μ_1	7%	A_5	7%
μ_6	9%	μ_2	7%	μ_2	7%
Kurt ₆	6%	μ_6	6%	μ_6^2	6%
μ_1^2	6%	μ_4	6%	μ_3^2	6%
μ_3^4	5%	Kurt ₆	6%	μ_1	5%
μ_6^3	4%	μ_4^3	4%	μ_5^2	5%
μ_3	4%	μ_4^2	4%	Kurt ₃	4%
μ_1^4	2%	μ_2^2	2%	μ_1^5	4%
μ_1	1%	μ_5	1%	A_4	3%
		Kurt ₅	1%	μ_3^5	3%
				μ_3^2	2%
				μ_4^4	1%
				μ_3^3	1%
				Kurt ₆	1%
				μ_6^3	2%
				Kurt ₃	1%
				μ_5^5	1%
				μ_1^3	1%
				μ_4	1%
				A_4	1%
				μ_1^4	1%
				μ_2^2	1%
				Kurt ₅	0.4%
				Kurt ₁	0.1%

Table 2 Normal Bayes classifier accuracy

	x	$x_{D4}^{0\%}$	$x_{D4}^{5\%}$	$x_{D4}^{10\%}$	$x_{D9}^{0\%}$	$x_{D9}^{5\%}$	$x_{D9}^{10\%}$	$x_{D49}^{0\%}$	$x_{D49}^{5\%}$	$x_{D49}^{10\%}$	x_μ	x_V	$x_{\mu+V}$
<i>D4</i>	0.78	0.87	0.88	0.68	0.87	0.88	0.68	0.87	0.88	0.68	0.90	0.68	0.91
<i>D9</i>	0.85	0.86	0.91	0.79	0.86	0.91	0.79	0.86	0.91	0.79	0.92	0.69	0.90
<i>D49</i>	0.85	0.88	0.91	0.60	0.88	0.91	0.60	0.88	0.91	0.60	0.90	0.49	0.91

Table 3 CART classifier accuracy

	x	$x_{D4}^{0\%}$	$x_{D4}^{5\%}$	$x_{D4}^{10\%}$	$x_{D9}^{0\%}$	$x_{D9}^{5\%}$	$x_{D9}^{10\%}$	$x_{D49}^{0\%}$	$x_{D49}^{5\%}$	$x_{D49}^{10\%}$	x_μ	x_V	$x_{\mu+V}$
<i>D4</i>	0.87	0.86	0.82	0.54	0.86	0.82	0.54	0.86	0.81	0.54	0.76	0.75	0.82
<i>D9</i>	0.85	0.86	0.89	0.69	0.86	0.89	0.69	0.86	0.89	0.69	0.85	0.77	0.89
<i>D49</i>	0.80	0.80	0.82	0.59	0.80	0.82	0.59	0.80	0.82	0.59	0.79	0.74	0.82

3.2.3 Walking Speed Influence

In the next stage of the research the influence of the robot's walking speed on the classification accuracy was assessed. In order to do it two experiments were performed. Firstly, the $D4$ set was used as a training set and the $D9$ set was used as the testing set. Secondly, the $D9$ served as a training set while the $D4$ was used for testing. All the previously defined feature vectors were evaluated and both the NBC and the CART classifiers were used. The obtained classification accuracy is presented in the table 4. It is clearly visible that the change of walking speed led to significant accuracy deterioration. It is worth noting that when the training set contained samples recorded during different walking speeds the classification was accurate (see the subsection 3.2.1). Therefore, in order to obtain an accurate and robust classifier it is important to gather training samples for a broad range of possible walking speeds.

Table 4 Accuracy under changing walking speed

Type	Trg.	Test	x	$x_{D4}^{0\%}$	$x_{D4}^{5\%}$	$x_{D4}^{10\%}$	$x_{D9}^{0\%}$	$x_{D9}^{5\%}$	$x_{D9}^{10\%}$	$x_{D49}^{0\%}$	$x_{D49}^{5\%}$	$x_{D49}^{10\%}$	x_μ	x_V	$x_{\mu+V}$	
NBC	$D4$	$D9$		0.11	0.14	0.17	0.17	0.14	0.17	0.17	0.14	0.17	0.17	0.15	0.10	0.19
NBC	$D9$	$D4$		0.20	0.21	0.22	0.18	0.21	0.22	0.18	0.21	0.22	0.18	0.22	0.19	0.23
CART	$D4$	$D9$		0.09	0.09	0.16	0.17	0.09	0.16	0.17	0.089	0.16	0.17	0.11	0.12	0.15
CART	$D9$	$D4$		0.09	0.10	0.14	0.10	0.10	0.14	0.10	0.10	0.14	0.10	0.14	0.12	0.17

3.2.4 Signal Length Influence

In the final experiment the robustness of the classification with regard to the length of the registered signals and the synchronization of the registration procedure with the robot's step was assessed. The $D9$ set was used in the experiment. The signal lengths of $n \in \langle 10, 20, \dots, 900 \rangle$ were analyzed. 20 trials were executed for each of the signal lengths and the average accuracy for each length was calculated. In each trial a new dataset was created by selecting subvectors of length n from each of the $D9$ elements:

$$s_i^n = s_i(m : m + n - 1) \quad (2)$$

where s_i^n is the created i -th subvector, s_i is the i -th element of the $D9$ set and m is a random integer from the range $\langle 1, 900 - n \rangle$. The leave-one-out approach was used to evaluate the accuracy of classifiers for each set.

The figures 4 and 5 present the obtained average accuracy as a function of signal length. It can be clearly seen that even significantly shorter signals registered asynchronously of the robot's steps can be successfully used to recognize the terrain type.

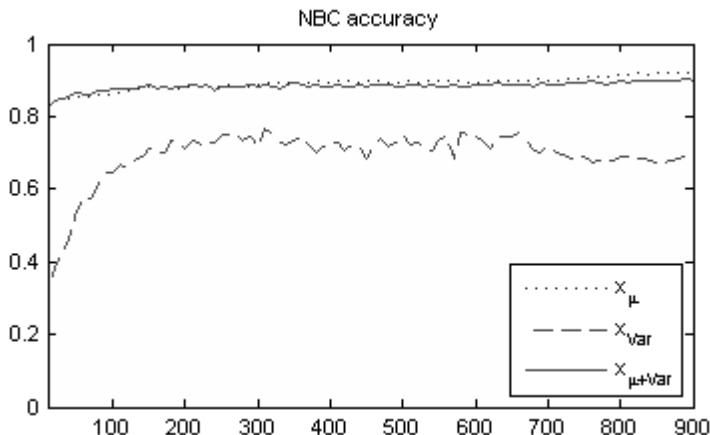


Fig. 4 The influence of signal length on the NBC accuracy

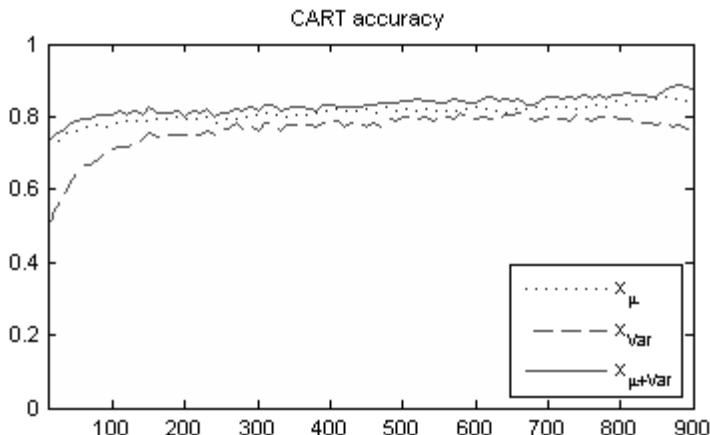


Fig. 5 The influence of signal length on the CART accuracy

4 Conclusions and Future Work

This paper presented a terrain type recognition system for a hexapod robot. It has been shown that the means and variances of force/torque signals and simple classifiers such as the Normal Bayes Classifier or CART are sufficient to achieve over 90% accuracy. It is necessary to include training samples registered at different walking speeds in the training set in order to achieve robustness w.r.t. the robot's walking speed. Moreover, the proposed system achieves high accuracy even for the signals registered only over small part of the robot's step.

In the future different classification approaches such as support vector machines and tree ensembles will be evaluated. Moreover, different feature extraction approaches such as Fourier analysis or wavelet decomposition will be used. Additionally, a dedicated hardware recognition system will be designed and implemented in the FPGA.

References

1. Bishop, C.M.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus (2006)
2. Breiman, L., Friedman, J., Olshen, R., Stone, C.: *Classification and Regression Trees*. Wadsworth and Brooks, Monterey (1984)
3. Brooks, C., Iagnemma, K.: Vibration-based terrain classification for planetary exploration rovers. *IEEE Transactions on Robotics* 21(6), 1185–1191 (2005)
4. Filitchkin, P., Byl, K.: Feature-based terrain classification for LittleDog. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1387–1392 (2012)
5. Giguère, P., Dudek, G., Saunderon, S., Prahacs, C.: Environment Identification for a Running Robot Using Inertial and Actuator Cues. In: Sukhatme, G.S., Schaal, S., Burgard, W., Fox, D. (eds.) *Robotics: Science and Systems*. The MIT Press (2006)
6. Giguere, P., Dudek, G.: Clustering sensor data for autonomous terrain identification using time-dependency. *Auton. Robots* 26(2-3), 171–186 (2009)
7. Giguere, P., Dudek, G.: Surface identification using simple contact dynamics for mobile robots. In: *IEEE International Conference on Robotics and Automation, ICRA 2009*, pp. 3301–3306 (May 2009)
8. Giguere, P., Dudek, G.: A Simple Tactile Probe for Surface Identification by Mobile Robots. *IEEE Transactions on Robotics* 27(3), 534–544 (2011)
9. Hoepflinger, M., Remy, C., Hutter, M., Spinello, L., Siegwart, R.: Haptic terrain classification for legged robots. In: *2010 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2828–2833 (2010)
10. Hoffman, R., Krotkov, E.: Perception of rugged terrain for a walking robot: true confessions and new directions. In: *Proceedings of the IEEE/RSJ International Workshop on Intelligent Robots and Systems 1991, Intelligence for Mechanical Systems, IROS 1991*, vol. 3, pp. 1505–1510 (November 1991)
11. Krotkov, E.: Active perception of material and shape by a walking robot. In: *Fifth International Conference on Advanced Robotics, 1991, Robots in Unstructured Environments, ICAR 1991*, vol. 1, pp. 37–42 (1991)
12. Ojeda, L., Borenstein, J., Witus, G., Karlsen, R.: Terrain characterization and classification with a mobile robot. *Journal of Field Robotics* 23(2), 103–122 (2006)
13. Posner, I., Cummins, M., Newman, P.: A generative framework for fast urban labeling using spatial and temporal context. *Autonomous Robots* 26(2-3), 153–170 (2009)
14. Walas, K., Belter, D.: Messor – versatile walking robot for search and rescue missions. *Journal of Automation, Mobile Robotics & Intelligent Systems* 5(2), 28–34 (2011)
15. Weiss, C., Frohlich, H., Zell, A.: Vibration based Terrain Classification Using Support Vector Machines. In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4429–4434 (October 2006)

Robust Registration of Kinect Range Data for Sensor Motion Estimation

Michał Nowicki and Piotr Skrzypczyński

Abstract. This work concerns the problem of determining the rototranslation between two 3D data sets. The sensor being used is Kinect, which yields large amount of data, thus processing all the point clouds in real-time on a standard PC is impossible. Therefore we analyse and compare two approaches: the standard ICP algorithm, and a method that uses salient point features to reduce the amount of data. To obtain a range data registration procedure, which is both precise and robust to large displacements of the sensor we combine these two methods.

1 Introduction

Determining the displacement of a sensor between two consecutive points on a trajectory is a classic problem in both computer vision and robotics. Known algorithms that compute the sensor's displacement from passive vision data are popular in robotics, but usually they require specific discrete features to be present in the environment or rich textures on the observed surfaces [4]. Also point clouds yielded by 3D laser scanners can be used to determine the displacement of a robot. This approach is relatively independent from particular geometric and photometric characteristics of the environment, because dense point clouds are matched against each other, rather than specific features. The procedure for data registration, i.e. for determining sensor displacement between consecutive points on the trajectory is typically implemented with the Iterative Closest Points (ICP) algorithm [1]. Nowadays this approach can be implemented not only with heavy and bulky 3D lasers with mechanical scanning, but also with much smaller and affordable 3D cameras, such as the Kinect sensor [15].

The solution proposed in this paper is based on the concept of matching consecutive point clouds using the ICP paradigm. However, it was found that the point-to-point matching of clouds yielded by Kinect requires a good initial esti-

Michał Nowicki · Piotr Skrzypczyński

Institute of Control and Information Engineering, Poznań University of Technology,
ul. Piotrowo 3A, 60-965 Poznań, Poland
e-mail: michalsminowicki@gmail.com,

piotr.skrzypczynski@put.poznan.pl

mate of the robot displacement. In a typical wheeled robot such an estimate can be obtained from odometry, but in tracked and walking robots or wearable devices proprioceptive sensing is unreliable, and thus dead reckoning does not exist or is extremely prone to errors. Therefore, feature-based matching of point clouds was implemented as an alternative method to obtain the initial displacement estimate. This matching procedure employs the NARF detectors/descriptors [14]. Implementation uses the open source Point Cloud Library (PCL) [12] to manipulate the 3D points and to detect the NARFs. Such a system can be applied to a broad class of small mobile robots [9], or used in wearable robotic devices [3].

2 Related Work

The Kinect sensor quickly got very popular in the computer vision and robotics community. Although Kinect was developed as a motion controller for computer games, it is used in a wide variety of robotics-related applications, due to its affordability and the possibility to obtain both the range data and the RGB images [7].

The 3D data sets obtained from the Kinect sensor are very large. Therefore, the ICP algorithm, which tries to establish correspondences between all the points being available, consumes a considerable time amount for each iteration. This, in most cases, makes impossible to estimate the sensor egomotion in real time. Such amount of data can be processed using a parallel implementation of ICP on GPGPU [10], but this type of hardware is not available for on-board processing in most of the contemporary robots. The KinectFusion system [5] is also an example of using the ICP algorithm on the range data. It was demonstrated that this method is enough for real-time tracking of the sensor with unconstrained motion, as the Kinect is used in this system like a handheld camera, without odometry. An implementation of KinectFusion is available within the Point Cloud Library, but this system heavily uses the GPGPU, and thus it is unavailable to many small and low-cost robotic systems.

Endres et al. [2] demonstrated recently a full SLAM system using Kinect RGB and depth information. This solution relies mostly on the photometric data (RGB images) for feature extraction, and then uses the depth image to locate the keypoints in 3D space. While this system shows good performance on large data sets, it requires small displacements between the consecutive frames. Therefore, the RGB-D images from Kinect have to be acquired and processed at the full speed, which is again impossible with limited computing power.

3 Range Data Registration

3.1 *The Basic Algorithm for Point Cloud Matching*

The ICP algorithm in its basic version was proposed in [1]. In each iteration, the ICP algorithm selects the closest points (Euclidean distance) in both clouds as the

matching points. For matching points, the rotation \mathbf{R} and translation \mathbf{t} are computed, which minimize the criteria:

$$(\mathbf{R}, \mathbf{t}) = \underset{\mathbf{R}, \mathbf{t}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} w_{i,j} \| \mathbf{p}_{A_i} - (\mathbf{R}\mathbf{p}_{B_j} + \mathbf{t}) \|^2 \right\}, \quad (1)$$

where n_A and n_B are the numbers of points in the clouds \mathcal{A} and \mathcal{B} , respectively, $w_{i,j}$ are weights defined as $w_{i,j} = 1$ when points \mathbf{p}_{A_i} and \mathbf{p}_{B_j} are closest neighbors or $w_{i,j} = 0$ otherwise. Assuming that points are matched correctly, the transformation (\mathbf{R}, \mathbf{t}) is computed by using the Singular Value Decomposition.

Once the transformation that minimizes (1) is determined, the \mathcal{A} data set is transformed to the new position and new matchings of points are established. Due to the search for the closest neighbors the computational complexity of ICP is quadratic with regard to the number of points in the two clouds. A more effective way of searching neighbors can be achieved using kd-trees to represent the data. Because the clouds may not represent the same part of the scene, some restrictions for the maximum Euclidean distance between neighbors must be introduced. These restrictions, expressed by the d_{\max} distance, prevent the algorithm from matching points that are too far away from each other. The ICP algorithm is only guaranteed to converge to a local minima, and may not reach the global one [1]. Because of this, it is extremely important to have a good initial guess for the sought (\mathbf{R}, \mathbf{t}) transformation.

In the literature, one may find different variants and improvements of the ICP. They involve speeding up the algorithm [11], improving its robustness to noise in range data, and improving robustness to wrong initial guess of the transformation [13]. In the research described in this paper the implementation of ICP that is available in the PCL library was used. Although some modifications of this algorithm, like the point-to-plane ICP, are considered better than vanilla ICP in the literature, the choice of the standard algorithm makes it possible to asses how much the proposed feature-based matching method improves the data registration with regard to the baseline method.

3.2 Matching Point Clouds Using Features

An alternative approach to matching data exploits salient features of different types. This approach is similar to matching 2D images from video cameras [8]. It is based on finding some keypoints in both data sets, then describing them with unique descriptors, and finding the matching between the computed descriptors. It is needed to mention, that these features do not represent pre-specified geometric structures (edges, corners), and commonly appear in both structured (man-made) and unstructured environments.

As we are interested in the range data from Kinect, which we consider to be a richer representation of the environment than the photometric (RGB) data, we choose the Normal Aligned Radial Feature (NARF) detector/descriptor for our implementation of the feature-based point cloud matching procedure. The NARF

concept was recently introduced for object recognition from 3D range data [14]. The NARF detector searches for stable areas with significant changes in the neighborhood, that can be keypoints looking similar from various viewpoints. Then, the NARF descriptors are built upon the areas around the keypoints by calculating a normal aligned range value patch, and identifying the dominant orientation of the points around.

The most important parameter of the NARF detector is the diameter of the sphere around the processed point. This sphere contains points, which dominant directions are used to determine the “coefficient of interest”, and then to detect the keypoints. In the detected keypoints the NARF descriptors are computed. Those descriptors are the normalized values of gradient in n directions with regard to the keypoint. The number of directions is also the length of the NARF descriptor (in this paper $n=36$).

3.3 Estimating the Transformation

In the keypoint matching algorithm the input data consists of a point set \mathcal{A} (the first cloud) and a point set \mathcal{B} (the second cloud). The set of all possible pairs of descriptors is defined as \mathcal{Z} . For each descriptor of the first point cloud all possible pairings with the descriptors of the second cloud are established. For the set \mathcal{Z} the matchings are defined by the descriptor dissimilarity function, which in our application is given as:

$$f_k = \min_{j=0,1,\dots,m_B-1} \sum_{i=1}^{36} |p_k[i] - q_j[i]|, \quad (2)$$

where f_k is the value of dissimilarity, which is the sum of absolute differences between 36 values describing the k -th descriptor of the cloud \mathcal{A} , and the corresponding values of the descriptor of the cloud \mathcal{B} . The descriptor of the cloud \mathcal{B} is chosen to minimize the value of f_k . In equation (2) $p_k[i]$ is the i -th value of the k -th descriptor of the first point cloud, while $q_j[i]$ is the i -th value of the j -th descriptor of the second point cloud, and m_B is the number of descriptors of the second cloud. Then, all pairings of the given descriptor of \mathcal{A} are investigated, and the pairs for which the dissimilarity value is greater than 3 times f_k are discarded as incorrect.

Among the accepted pairings there could be some incorrect correspondences, which lead to a clearly incorrect measurement of the displacement. However, it is possible to introduce a set of constraints on the maximum translations and rotations with regard to the particular axes of the sensor’s coordinate frame. These constraints are used to discard pairs of the descriptors for which the distance between their keypoints is bigger than some maximum Euclidean distance.

Because the set of matching descriptors still can contain some outliers, a robust framework based on the RANSAC scheme is introduced to estimate the transformation. In this framework the RANSAC procedure in each of its iterations defines a subset \mathcal{N} of the set \mathcal{Z} , which contains 3 unique descriptor pairs. The next step of the algorithm computes the transformation between those pairs:

$$\underset{\mathbf{T}}{\operatorname{argmin}} \|\mathbf{TP} - \mathbf{Q}\|, \quad (3)$$

where \mathbf{T} means the transformation combined of rotation \mathbf{R} and translation \mathbf{t} , while \mathbf{P} and \mathbf{Q} are matrices built from the coordinates of keypoints as follows:

$$\mathbf{P} = \begin{bmatrix} x_{11} & y_{11} & z_{11} \\ x_{12} & y_{12} & z_{12} \\ x_{13} & y_{13} & z_{13} \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} x_{21} & y_{21} & z_{21} \\ x_{22} & y_{22} & z_{22} \\ x_{23} & y_{23} & z_{23} \end{bmatrix}, \quad (4)$$

where x_{1i}, y_{1i}, z_{1i} are the coordinates of the i -th keypoint from the cloud \mathcal{A} , belonging to the set \mathcal{N} , while x_{2j}, y_{2j}, z_{2j} are the coordinates of the j -th keypoint from the cloud \mathcal{B} , also belonging to the set \mathcal{N} . To solve the problem stated as (3) (known as the Procrustes Problem) the simple and fast Kabsch algorithm is used. After computing the transformation \mathbf{T} new coordinates for the points of the cloud \mathcal{A} are calculated. The next step involves computing the matching error:

$$\varepsilon = \sum_{i=1}^{n_A} (\mathbf{h}_i - \mathbf{d}_j)^2 \quad \text{for } j = \underset{j=0,1,\dots,n_B-1}{\operatorname{argmin}} (\mathbf{h}_i - \mathbf{d}_j)^2, \quad (5)$$

where \mathbf{h}_i is the vector of coordinates of i -th keypoint of the point cloud \mathcal{A} transformed by \mathbf{T} , \mathbf{d}_j is the vector of coordinates of the j -th keypoint of the cloud \mathcal{B} , which is the closest to \mathbf{h}_i . When the value of $(\mathbf{h}_i - \mathbf{d}_j)^2$ is greater than a fixed parameter, the keypoint is treated as an outlier and the computed distance is not counted in (5).

The condition of stopping RANSAC is achieving the error ε below a preset value, or not improving the best model for a fixed number of iterations, or exceeding the number of maximum iterations. When RANSAC is finished, the transformation with minimum error is performed on all points of the cloud \mathcal{A} . This algorithm allows to achieve a good starting transformation for the ICP algorithm. In order to transform the point cloud \mathcal{A} to the starting point, the following must be computed:

$$\mathbf{p}'_A[i] = \mathbf{R}\mathbf{p}_A[i] + \mathbf{t} \quad \text{for } i = 1, 2, \dots, n_A \quad (6)$$

where $\mathbf{p}_A[i]$ means the i -th point of the first point cloud. The modified set \mathcal{A}' is then the point cloud for the ICP algorithm searching for a more precise transformation of \mathcal{A}' into \mathcal{B} .

3.4 Implementation of the Data Registration Method

The whole procedure of data registration is implemented in several steps (Fig. 1). At first, the range data are trimmed to discard the measurements shorter than 50 cm and longer than 350 cm, that is unreliable or of low depth resolution. Then, the VoxelGrid filter from the PCL library is used on both data sets with the voxel size of 2 cm. This filter reduces the number of points and decreases the influence of

small range measurement errors. We do not use other filters available in the PCL, as their use did not improve the results, while it consumed additional time for processing. NARF keypoints are detected in both point clouds under consideration, and then their descriptors are computed. The next step is matching of the descriptors as described in section 3.2. If a sufficient matching is found, the transformation between the two point clouds is estimated, and the data set \mathcal{A} is transformed to the new coordinates. The kinematic constraints are imposed in this step. Then, the ICP procedure from the PCL library is applied to the two data sets in order to obtain a precise transformation.

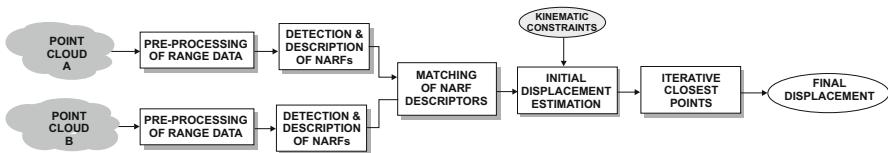


Fig. 1 Block scheme of the data registration procedure

4 Experiments and Results

4.1 Tests in a Controlled Environment

Preliminary experiments were carried out in a controlled environment with the use of an industrial robot KUKA KR-200. The high precision of the industrial arm allowed us to obtain ground truth for the motion of the sensor (Fig. 1A). Two point clouds collected during the experiment are shown in Fig. 1B – the sensor moved 20 cm along a straight line between these two poses. The cloud \mathcal{A} is shown in red, while the cloud \mathcal{B} in green. Small squares mark NARF keypoints for each of the clouds. Descriptors, used to determine the transformation between the clouds, are marked with larger squares.



Fig. 1 Terrain mockup (A), point clouds with NARFs prior to matching (B), matching result NARF+ICP (C)

In this experiment the standard ICP algorithm was compared to our NARF+ICP approach, in which the initial displacement estimate is computed from matching of

NARFs, as described in section 3.2. The ICP algorithm applied directly to the clouds from the Kinect was tested for different values of the d_{\max} parameter, but it turned out to be impossible to find a sensor displacement that was similar to the actual one. However, the feature-based matching algorithm was able to find a good initial transformation for the ICP algorithm. The number of accepted descriptor pairs was 58, and the number of RANSAC iterations reached 100. As a result of the whole data registration algorithm, i.e. NARF+ICP the displacement of 17.3 cm was obtained (Fig. 1C). In the whole data registration procedure the detection and description of the NARFs took 0.84 s, matching of NARFS including the RANSAC procedure required 0.25 s, and then the ICP algorithm ran for 0.27 s. The time was measured under Linux on a single-core low-end PC notebook.

4.2 Application-Oriented Experiments

Application-oriented experiments were conducted on the Messor walking robot (Fig. 2A) The robot was walking on a level floor along a straight line towards a group of objects, which are visible on the rendered RGB-D image (Fig. 2B). The knowledge of the motion direction and the estimated length of the step were used to impose constraints on the possible rotations and translations computed by the data registration algorithm. It should be noted, that the constraints used for data registration are not so tight: 50 cm translation along each axis, and 30° rotation around each axis are allowed. These values are not violated even by sudden and unpredictable motions related to slippages of the walking robot [9].

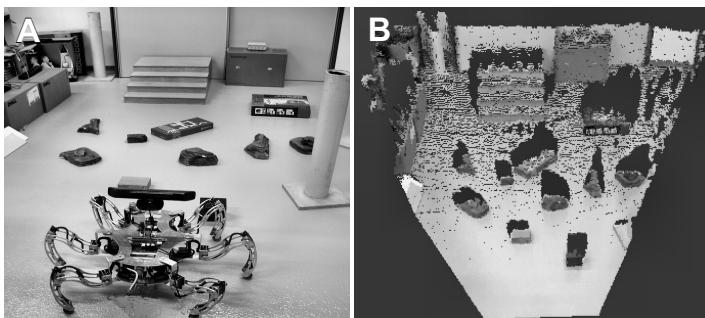


Fig. 2 Kinect mounted on our robot (A), and the indoor scene rendered from the RGB-D data (B)

The data set was obtained for a pure translational motion of 20 cm along the robot's y axis (Fig. 3A). The first experiment with this data set was carried out for the basic ICP algorithm. In spite of increasing the number of iterations and making trials with various settings of the d_{\max} parameter, it turned out that the solution contained rotations and translations in all axes, what is clearly wrong (Fig. 3B).

Table 1 Comparison of results for the indoor experiments – 20 cm translation

	kinematic constraints 1	no constraints
number of descriptors in \mathcal{A}	46	38
number of descriptors in \mathcal{B}	47	29
number of pairs	266	765
RANSAC iterations	939	93669
duration (all steps but ICP)	1.32 s	23.5 s
rotation θ_r	1.08°	-40.25°
rotation ϕ_r	1.13°	0.54°
rotation ψ_r	1.13°	1.79°
translation in x	2.2 cm	59 cm
translation in y	7.0 cm	80.8 cm
translation in z	20.5 cm	28.9 cm

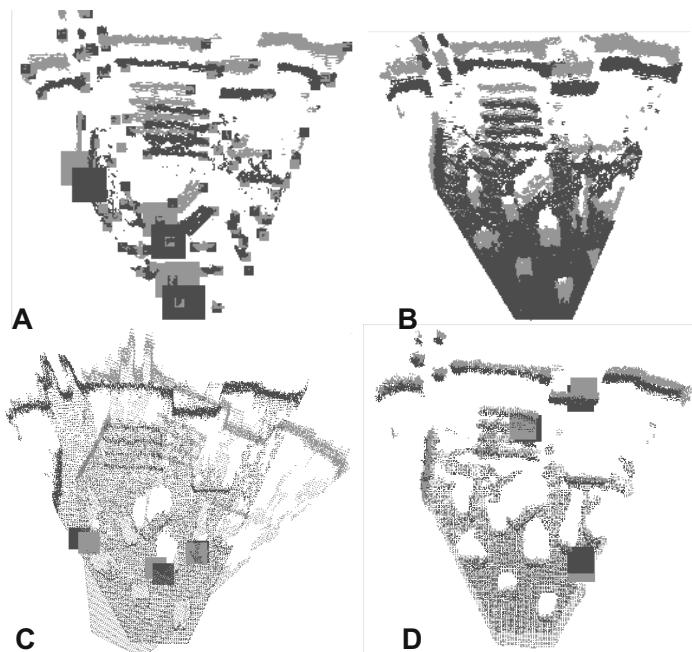
**Fig. 3** Indoor experiment – two point clouds prior to matching (A), results of matching with ICP (B), results of matching with NARF+ICP (C), results of matching with NARF+ICP and constraints imposed (D)

Table 1 provides sample displacement estimation results for two versions of the data registration procedure: the one with kinematic constraints resulting from the known robot control commands, and the one in which these constraints are not

imposed. When the constraints are not used the found transformation is inferior to the one found imposing the constraints, in spite of the much longer time taken by the entire procedure and the large number of RANSAC iterations (Fig. 3C).

The matching procedure used the RANSAC scheme to discard false matches, and the kinematic constraints were applied to discriminate wrong pairings of descriptors. Owing to this processing scheme the procedure was robust enough to produce a sufficient initial guess of the transformation between the two clouds: 15.3 cm. Starting from this value the ICP algorithm produced a more precise estimate of the translation: 20.5 cm (Fig. 3D). In this case the NARF detection and description procedure took 1.03 s, because of the large number of keypoints. However, matching with RANSAC took 0.29 s, and the ICP method required 0.40 s, that is the speed was similar as in the controlled environment. The walking robot needed about four seconds to cover the 20 cm distance, using a tripod gait and making steps of the length of 5 cm. Thus, the point cloud registration procedure can be described as real-time, as the low-end PC was able to complete all registration steps between the two consecutive moments of data acquisition.

5 Conclusions

The experiments have shown that the use of NARF descriptor matching on the point clouds from Kinect allows to obtain a good initial estimate for the ICP algorithm even for relatively large displacements. Therefore the developed robust registration procedure may be applied to point clouds obtained sparsely on the trajectory of a robot. Because the procedure applies only very general kinematics constraints it can be used to any type of motion in 2D or 3D, and thus applied to many types of robots. Also the moderate requirements as to the computing power (a standard low-end notebook is enough) make it possible to apply our method to virtually any robotic device, including walking, humanoid and wearable robots.

References

1. Besl, P.J., McKay, N.D.: A Method for Registration of 3-D Shapes. *IEEE Trans. on Pattern Analysis and Machine Intell.* 14(2), 239–256 (1992)
2. Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., Burgard, W.: An Evaluation of the RGB-D SLAM System. In: Proc. IEEE Int. Conf. on Robot. and Automat., pp. 1691–1696. St. Paul (2012)
3. Fallon, M., Johannsson, H., Brookshire, J., Teller, S., Leonard, J.: Sensor Fusion for Flexible Human-Portable Building-Scale Mapping. In: Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Systems, Vilamoura, pp. 4405–4412 (2012)
4. Fraundorfer, F., Scaramuzza, D.: Visual Odometry: Part II - Matching, Robustness and Applications. *IEEE Robotics and Automation Magazine* 19(2), 78–90 (2012)

5. Izadi, S.: KinectFusion: Real-Time 3D Reconstruction and Interaction Using a Moving Depth Camera. In: ACM Symp. on User Interface Software and Technology, New York, pp. 559–568 (2011)
6. Kabsch, W.: A Solution of the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallographica* 32, 922 (1976)
7. The Kinect Effect – How the World is Using Kinect (2012),
<http://www.xbox.com/en-GB/kinect/kinect-effect>
8. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *Int. Journal of Computer Vision* 60(2), 91–110 (2004)
9. Nowicki, M., Skrzypczyński, P.: Experimental Verification of a Walking Robot Self-Localization System with the Kinect Sensor, *Prace Naukowe Politechniki Warszawskiej – Elektronika. Postępy robotyki*, Warsaw, pp. 561–572 (2012) (in Polish)
10. Nüchter, N., Feyzabadi, S., Qiu, D., May, S.: SLAM à la Carte – GPGPU for Globally Consistent Scan Matching. In: Proc. 5th European Conf. on Mobile Robots (ECMR), Örebro, pp. 271–276 (2011)
11. Rusinkiewicz, S., Levoy, M.: Eficient Variants of the ICP Algorithm. In: Proc. 3rd Int. Conf. on 3D Digital Imaging and Modeling, Quebec, pp. 145–152 (2001)
12. Rusu, R.B., Cousins, S.: 3D is Here: Point Cloud Library (PCL). In: Proc. IEEE Int. Conf. on Robot. and Automat., Shanghai, pp. 1–4 (2011)
13. Segal, A., Haehnel, D., Thrun, S.: Generalized-ICP. In: Proc. of Robotics: Science and Systems, Seattle (2009)
14. Steder, B., Rusu, R.B., Konolige, K., Burgard, W.: Point Feature Extraction on 3D Range Scans Taking Into Account Object Boundaries. In: Proc. IEEE Int. Conf. on Robot. and Automat., Shanghai, pp. 2601–2608 (2011)
15. Stoyanov, T., Louloudi, A., Andreasson, H., Lilienthal, A.: Comparative Evaluation of Range Sensor Accuracy in Indoor Environments. In: Proc. 5th European Conf. on Mobile Robots (ECMR), Örebro, pp. 19–24 (2011)

Utilization of Depth and Color Information in Mobile Robotics

Maciej Stefańczyk, Konrad Bojar, and Włodzimierz Kasprzak

Abstract. Computer vision plays an increasing role in robotics, as the computing power of modern computers grows year by year allowing more advanced algorithms to be implemented. Along with visual information, depth is also widely used in navigation of mobile robots, for example for obstacle detection. As cheap depth sensors become popular nowadays, there is a possibility to use data from both sources to further enhance the processes of navigation and object detection. This article presents some possibilities of utilizing in mobile robotics the integrated video and depth images - by performing image segmentation for environment description, optical flow estimation for obstacle avoidance and object detection for semantic map creation. All of the presented examples are based on real, working applications, which additionally proves validity of proposed methods.

1 Introduction

In recent years, thanks to advances in computer entertainment systems, sensors capable of acquiring three dimensional scene data became cheap and popular. The Microsoft Kinect device (created to be an advanced motion sensor for the XBox console) was adopted in just a few days after release (in 2010) to work with PC computers [4], and from the very beginning found its place in computer vision, robotics and cognitive systems all around the world. Obviously, depth sensing was used in robotics long before Kinect and alike devices were popular, but in most of applications there was a strict border between the use of depth and color information

Maciej Stefańczyk · Włodzimierz Kasprzak
Institute of Control and Computation Engineering, Warsaw University of Technology,
Nowowiejska 15/19, 00-665 Warsaw, Poland
e-mail: {M.Stefanczyk,W.Kasprzak}@elka.pw.edu.pl

Konrad Bojar
Industrial Research Institute for Automation and Measurements, Al. Jerozolimskie 202,
02-486 Warsaw, Poland
e-mail: kbojar@piap.pl

- the first one was used only for obstacle avoidance and localization, the latter one
- mostly for object detection. There are only some algorithms incorporating both sensors, those are mostly focused on visual odometry [14] or SLAM [3].

This article focuses on possible benefits of using depth information with associated color information in robotics, especially for mobile platforms. At first, in section 2 some background information and motivation is given, followed by a short survey of existing, easy to use methods of associated depth and color information acquisition (section 3). In subsequent sections (4-6) some robotic applications are presented, covering a variety of tasks, each gaining benefits from incorporating depth information in data processing.

2 Motivation

Out of available sensors, computer vision seems to be the most appealing to use in robotic applications, due to large amount of information it provides. Its application domain ranges from visual odometry [11], through self-localization using either natural [8] or artificial landmarks [1], to object detection [7]. Unfortunately, those readings are hard to process and understand if one needs something more complicated to be calculated in realtime. Image processing causes heavy computational load and control system has to be powerful enough to handle it. However, in recent years computational power of modern PCs has increased, and today one can equip almost every robot (even a small one) with a relatively powerful and cheap computer. Computer vision plays a main role in almost every application, where interaction with environment is a crucial need. Machine-human interaction, automated guide robots in museums and service robots in hospitals or autonomous terrain vehicles are just a few examples from wide variety of applications that require a visual perception ability.

However, vision itself is sometimes still not enough to handle real-world problems. Those can be caused either by sensor errors or acquisition methods imperfections, like limited dynamic range resulting in under- or overexposed parts in image, or by specific algorithm disadvantages, like inability to distinguish between many plain, textureless objects. Some of those problems could be solved by using an active-vision paradigm, but this leads to complicated control systems. In some applications relying only on color image gathered by cameras is simply not sufficient. When the task assumes grasping of unknown objects, their accurate 3D model has to be built in realtime, which is much easier to accomplish when depth information is available.

3 Sources of Depth Maps with Registered Color Information

Distance sensors were used in robotics from years. Those include infrared and ultrasonic range detectors, as well as more advanced lidar solutions. That kind of sensors

can be easily incorporated into the navigation system of a mobile robot in order to detect obstacles, to create a terrain map or to handle emergency situations. Unfortunately, to acquire a complete depth map of a scene additional hardware is required. For lidar, tilting head is a must [15] and depending on control accuracy different vertical resolutions could be achieved. Sequential scanning requires long processing time to gather all readings and a complete map is created only once per every few seconds. Another problem is aligning depth images with color information, as both images are obtained from separate devices.

Another solution is a time-of-flight camera, where, in simple terms, the scene is illuminated with a strobe of light sequence and a sensor measures the time delay elapsed between flash time and returned light registration. This technology is much faster than previously mentioned solutions, reaching sometimes a frequency higher than 100 Hz, and in many implementations it is possible to extract gray scale images from the same sensor, which produces depth and video data being perfectly aligned in space. On the other hand, the TOF cameras create rather low resolution images (in general the pixel number is lower than 0.04 Mpix) and without color information, which makes this solution almost useless from the point of view of scene analysis. Anyway, a simple obstacle detection [12] is a good example of an application utilizing this data source.

Different approaches to depth map acquisition depend heavily on image analysis. The most popular among them are stereovision and structured light. Stereovision relies on images obtained from synchronized cameras (usually a pair), in which some characteristic features are localized and matched together. There exist hardware solutions with built-in processing units, that produce depth maps automatically, but most applications rely on software implementations, which are rather slow. The biggest problem with stereovision, when used in indoor robotics applications, is the sparsity of resulting depth map as office-like environments have only few characteristic points needed by a matching algorithm, which leads to partial and poor results (left part in fig. 1). There are solutions solving this problem, like projecting a texture onto the scene [9]. This method is widely used in outdoor applications, where almost a full depth coverage can be achieved in practice [10]. The main advantage of stereovision is that the resulting depth map is perfectly aligned with the color image, because it's calculated directly from it.

Another method of depth map acquisition based on image processing is the structured light approach. In its basic configuration there is one camera and an image projector, synchronized to each other (like in stereo vision). A scene pattern deformation seen by the camera gives information about the distance to a particular scene region. Simplest methods of this type use a line projector to model the ground shape [16], which is rather a low-demanding problem in terms of computing power. For restoration of a full frame information other patterns are needed – some algorithms use multiple patterns for single frame to recover depth information [6], others can produce depth map by a single exposure. To achieve this, rather complicated patterns and algorithms are required. In [2] authors propose color encoded patterns to speed-up the 3D-coordinates recovery, but this greatly influences normal camera readings and may be cumbersome for humans staying near to the robot. To make

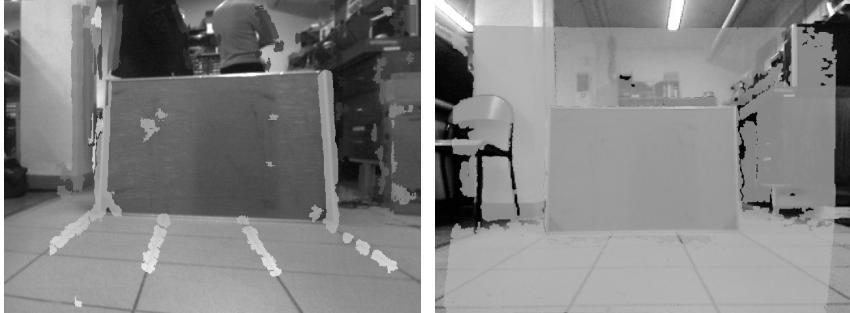


Fig. 1 Example results produced by stereovision (left) and structured light sensors (Kinect, right). The depth value encoded in rainbow-like palette is aligned with and overlaid onto the RGB image

structured light much less irritating for people, solutions working with light in near infrared band are used, with Microsoft Kienct being the widest known at present time. Using the IR projector one needs two different cameras – one for registering depth and the other for the color. This produces some required translation between them, and the resulting data is usually not perfectly-aligned and contains some shadowing effects. But for robotics purposes it constitutes a great practical solution, as it is working fast (works stable with the frequency of 30Hz) and it is extremely cheap (its price is about 100\$). The whole image processing is done by a dedicated hardware, leaving the CPU time to be spent on the processing of high amount of data it produces. On the right side of fig. 1 the output from Kinect is shown (for comparison with stereo vision). It's clearly visible, that stereo vision is much better in detecting narrow objects, like chair legs, that are completely ignored by Kinect (no reading at those points, as object is to small to recognize a pattern on it). The depth coverage is much better done by Kinect, especially on flat table in the center of the frame. The biggest drawback of this solution is almost complete disability to work in outdoor scenes, where ambient and direct sun lighting is much stronger than built-in IR projector.

4 Aiding Segmentation Process with Additional Data

Following sections present three different aspects of using depth information in robotic applications.

The first one is the support of a color image segmentation process, by incorporating associated depth readings. Image segmentation is a crucial part of many algorithms in robotics. On mobile platforms it is a part of object detection and tracking, for example when searching for some objects in rooms and corridors. The depth

information can be incorporated into the image segmentation algorithm in at least two ways. Common step for both of them is registering depth with color information, which gives as a result an augmented image, in which each pixel has a color and can have scene coordinates given in some coordinate frame (or an information is provided, that the depth couldn't be computed).

From such kind of data more additional features can be computed – among them the normal vectors being the most popular one. As a result, we have a multi-modal image, which can be processed by extended segmentation algorithms, like in [13], presented in fig. 1. That kind of representation has an important advantage – it is robust in case of partial data loss. If there is a lack of depth information in some parts of the image, then a standard, color-based segmentation is used for them locally. If for some reason, like underexposure of image, the color image lacks information or has a low dynamic range, then the depth may be sufficient to continue the segmentation process, by taking into account only geometric attributes.

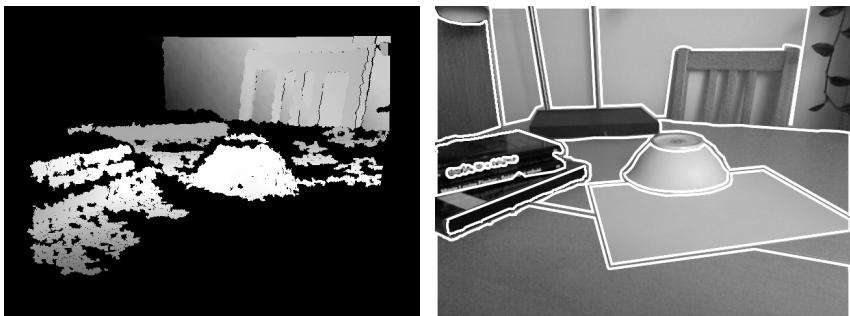


Fig. 1 Example of multi-modal segmentation results (from [13]). Such a segmentation process allows to recognize objects "not visible" at first view (e.g. a bowl on white paper). Smallest segments were filtered out for clarity.

Another approach is to use depth information in a filtering stage to remove parts of image that don't contain interesting information. If the algorithm searches for elevator doors for example, then the only interesting parts are those which are of vertical elongation. This filtering can be done using normal vectors stored for pixels. Raw depth information can be used to limit the search to a subspace of scene only - for example, when searching for balls laying on ground it is feasible to look at objects not taller than 5 cm. When the depth is unavailable, the detection algorithm have to cope with circular objects in the complete image, even if those are unreachable (e.g. placed too high). Fig. 2 shows an example scene with balls and the results of processing this scene with and without filtering that uses depth information. It is clearly visible, that some circles found only by color are rejected when depth is taken into account.

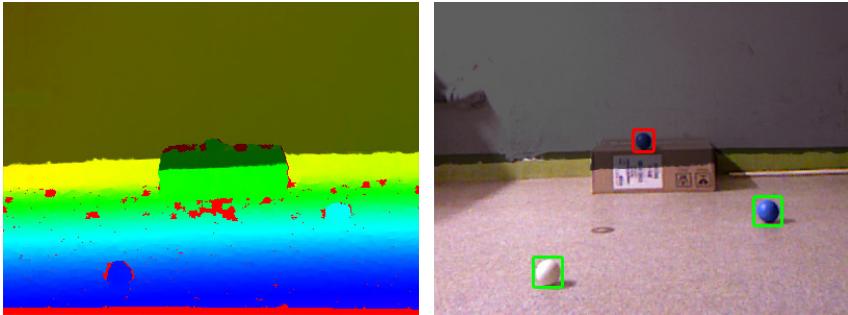


Fig. 2 Results of filtering image using depth information before color segmentation. Left image shows depth information, on right image reachable balls are marked green, unreachable is marked red. In both images darker area on upper part means unreachable zone (due to height)

5 Localization and Obstacle Detection by an Optical-Flow System

One of algorithms widely used in robotics is optical flow (OF) calculation [5], which can be used either for localization or obstacle detection. Although this algorithm itself can work using only visual data, when aligned depth is available we can easily recover full scene information, including real speed of objects around robot (with respect to robot or to some external coordinate frame). In this case, incorporating depth into the processing flow has many advantages over calculating OF on separate data types. If only color is used, then some additional computation is needed for determining 3D coordinates of vectors in image, and if only depth image is processed, although possible, can give poor results due to low amount of characteristic points (only edges are distinguishable). Aligned depth allows one to calculate movement vectors from color image, and then, knowing which pixels are transferred where, 3D coordinates of points can be simply read and motion vectors calculated.

Figure 1 presents optical flow (following original implementation proposed in [5]) calculated for a corridor scene with an obstacle located in front of the robot. Vectors are rather short as robot moves with low speed, but it's enough for determining real motion vectors from associated depth map. Each 2D vector from previous step is then transferred to a 3D vector M_i , producing a so called "scene flow" as a result. After removing vectors that are not a threat for the robot at the moment (neither being at robot height nor pointing in robots direction) information can be used as an input for a local planner that calculates repeatedly in time the repulsion force (1) applied to steering wheels.

$$F = \sum_i \text{sgn}(\alpha_i) \frac{\|M_i\|}{d_i} \cos(\alpha_i) \quad (1)$$

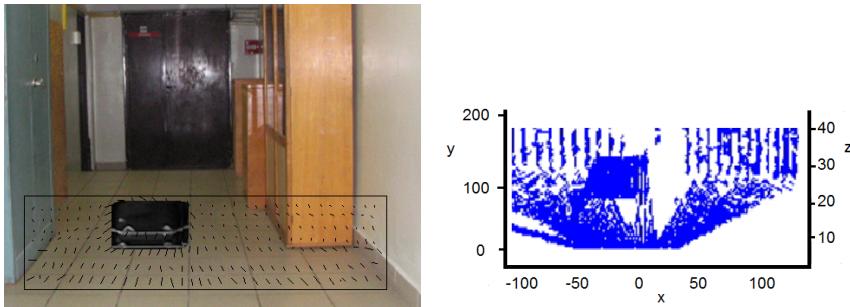


Fig. 1 The motion vectors extracted from the optical flow (left) are backprojected onto the scene with the help of the aligned depth map (right)

where d_i is distance between robot and i -th point in scene flow, α_i is angle between vector connecting robot with i -th point and motion vector M_i , and sign depends on side (relative to the robot) on which point is placed (negative if left, positive if right). Resulting value of F , after scaling, is sent to wheels (if repulsion is zero then robot goes forward, if negative then robot turns left, if positive then a right turn needs to be executed).

6 Verification Step in Object Detection

In section 4 a system detecting tennis balls was mentioned, which can benefit from the step of pre-filtering a color image. But this kind of system can still be tricked by placing a flat circle image on the floor, which would pass the depth filtering step and be detected. This kind of false readings would be of course rejected when robot approaches closer to the object, but it is much better to avoid unnecessary movements, especially in time critical applications.

To detect such false detections, each segment can be checked against the depth image. In the simplest case we expect corresponding features to be extracted in depth and in color images. For example, in the balls case in both images circular segments will exist in general. We can seek for some characteristic points in the color image (like SIFT) and check geometric properties in its surrounding area to estimate a 3D shape. If depth features are consistent with expectations, then a segment is passed to other modules. For example, such an object can play the role of a navigation goal for a motion planner.

The complete process of ball search and verification for a robot collecting tennis balls is as follows:

Searching:

1. pass the image through color filtering (remove dark and highly saturated objects),

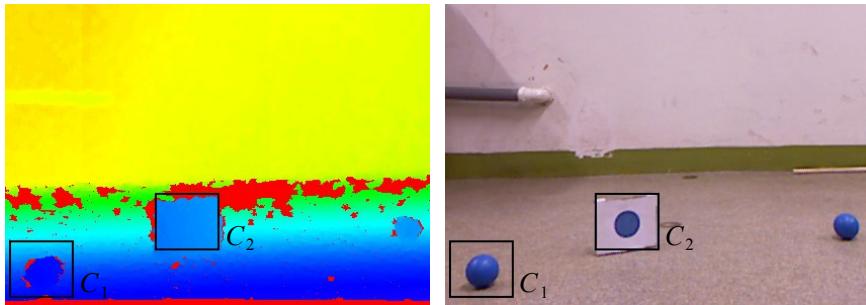


Fig. 1 Detecting tennis balls with verification using depth image

2. extract edges from resulting image,
3. use moment invariants to find circular objects C_i .

Verification:

1. for every detected object C_i extract its surrounding space from the depth image,
2. find the biggest circular object C_{ic} in extracted color image surrounding and extract edges C_{id} in the depth map,
3. if both objects are similar according to similarity ratio (2) then the object is assumed to be valid, otherwise it is rejected.

$$\text{sim}(C_{ic}, C_{id}) = \frac{P(C_{ic} \cap C_{id})}{P(C_{ic})} \quad (2)$$

where $P(C)$ means the number of pixels contained in a contour.

Figure 1 presents an example frame with results of the ball detection and verification algorithm. Two particular candidates are selected, one real ball and a ball

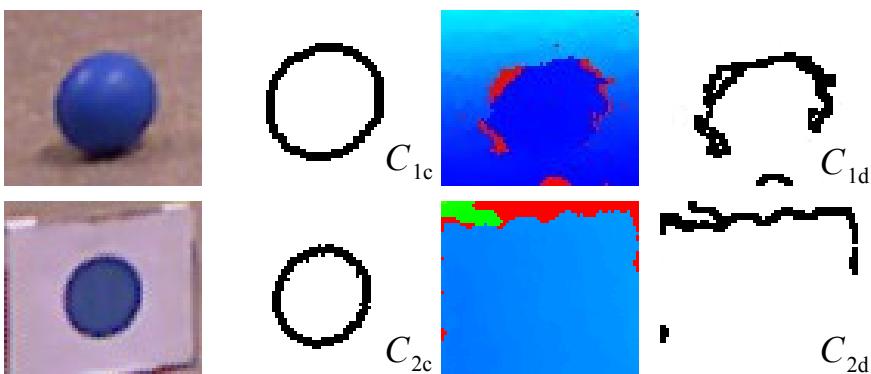


Fig. 2 Close-up of selected ball candidates from fig. 1

printed on a flat box. Those are candidate objects C_i , which close-ups are in figure 2. Those objects are processed according to presented steps. In case of C_1 , the contours extracted in the color image C_{1c} and from depth information C_{1d} are similar ($\text{sim}(C_{1c}, C_{1d}) = 0.42$), and the object is accepted to be a real ball. In the case of C_2 the similarity ratio $\text{sim}(C_{2c}, C_{2d}) = 0$, hence the object is rejected from further processing.

7 Summary

This paper presented several ways of incorporating depth readings that are aligned with a color image in order to improve robotic applications, especially those intended to be used on mobile platforms. As an introduction, available sources of 3D scene information were presented along with a short analysis and comparison of them. Three applications of such integrated depth and color images have been developed: environment map creation, motion-based obstacle avoidance and 3D object detection. The presented applications demonstrate the benefits of using additional depth information in typical color image analysis tasks. They lead to faster processing times and more accurate results of image analysis algorithms, while required changes of previously designed algorithms are small. Thus, we have verified that reliable depth information can be successfully used in existing computer vision applications.

Acknowledgements. The authors gratefully acknowledge the support of this work by The National Center for Research and Development, Warsaw - the grant PBS1/A3/8/2012.

References

1. Baczyk, R., Kasinski, A., Skrzypczynski, P.: Vision-based mobile robot localization with simple artificial landmarks. In: Artificial Landmarks, Prepr. 7th IFAC Symp. on Robot Control. Citeseer, Wroclaw (2003)
2. Chen, S.Y., Li, Y.F., Zhang, J.: Jianwei Zhang. Vision processing for realtime 3D data acquisition based on coded structured light. IEEE Transactions on Image Processing 17(2), 167–176 (2008)
3. Engelhard, N., Endres, F., Hess, J., Sturm, J., Burgard, W.: Real-time 3d visual slam with a hand-held rgb-d camera. In: Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden, vol. 2011 (2011)
4. Giles, J.: Inside the race to hack the Kinect. The New Scientist 208(2789), 22–23 (2010)
5. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial intelligence 17(1), 185–203 (1981)
6. Inokuchi, S., Sato, K., Matsuda, F.: Range imaging system for 3-d object recognition. In: Proceedings of the International Conference on Pattern Recognition, pp. 806–808 (1984)

7. Jung, B., Sukhatme, G.S.: Detecting moving objects using a single camera on a mobile robot in an outdoor environment. In: International Conference on Intelligent Autonomous Systems, pp. 980–987 (2004)
8. Kasprzak, W., Szymkiewicz, W.: A method for discrete self-localization using image analysis. In: Proceedings of the Third International Workshop on Robot Motion and Control, RoMoCo 2002, pp. 369–374. IEEE (2002)
9. Konolige, K.: Projected texture stereo. In: International Conference on Robotics and Automation (ICRA), pp. 148–155. IEEE (2010)
10. Konolige, K., Agrawal, M., Bolles, R.C., Cowan, C., Fischler, M.A., Gerkey, B.P.: Outdoor mapping and navigation using stereo vision. In: International Symposium on Experimental Robotics, pp. 179–190 (2006)
11. Nistér, D., Naroditsky, O., Bergen, J.: Visual odometry. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 1, pp. I–652. IEEE (2004)
12. Prusak, A., Melnychuk, O., Roth, H., Schiller, I., Koch, R.: Pose estimation and map building with a Time-Of-Flight-camera for robot navigation. *Int. J. Intell. Syst. Technol. Appl.* 5, 355–364 (2008)
13. Stefańczyk, M., Kasprzak, W.: Multimodal segmentation of dense depth maps and associated color information. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 626–632. Springer, Heidelberg (2012)
14. Steinbrucker, F., Sturm, J., Cremers, D.: Real-time visual odometry from dense rgbd images. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 719–722. IEEE (2011)
15. Surmann, H., Lingemann, K., Nüchter, A., Hertzberg, J.: A 3d laser range finder for autonomous mobile robots. In: 32nd International Symposium on Robotics (ISR), pp. 153–158 (2001)
16. Wei, B., Fan, Y., Gao, B.: Mobile robot vision system based on linear structured light and DSP. In: International Conference on Mechatronics and Automation, ICMA 2009, pp. 1285–1290 (August 2009)

Part VIII

Speech and Word Recognition

Texture-Based Text Detection in Digital Images with Wavelet Features and Support Vector Machines

Marcin Grzegorzek, Chen Li, Johann Raskatow,
Dietrich Paulus, and Natalia Vassilieva

Abstract. In this paper we propose to combine region-based and texture-based approaches for text detection in digital images. Our solution is based on a cascade filtering of image regions. First, we apply heuristic filtering to disregard certain non-textual areas. Second, we perform a more precise and expensive texture-based filtering using support vector machines and wavelet-based texture features. We have evaluated our approach with the ICDAR 2003 text locating competition benchmark collection and tools. The experimental results showed competitive performance of our solution by means of recall and precision compared to other text detection approaches participated in ICDAR 2003 and lower computational cost at the same time.

Keywords: Text Detection, Wavelet Features, Support Vector Machines.

1 Introduction

Text recognition in digital images is an important component for a wide range of applications such as automatic annotation and indexing of multimedia content, computerised aid for visually impaired, robotic navigation in urban environments and many others. Common OCR (Optical Character Recognition) tools are capable to transform document images into machine-readable text with a reasonably accuracy, but fail to detect text in images of natural scenes. However, it was shown that once being detected and localised, a text region in a scene image might be recognised by an OCR tool with a much better performance [2, 3, 15]. The reason is that a

Marcin Grzegorzek · Chen Li
Pattern Recognition Group, University of Siegen, Germany
<http://www.pr.informatik.uni-siegen.de>

Johann Raskatow · Dietrich Paulus
Active Vision Group, University of Koblenz-Landau, Germany

Natalia Vassilieva
HP Labs, St. Petersburg, Russia

background inside a localised text region is usually less complex compared to the whole image, and characters can easier be recognised. Another reason to perform text detection as a separate task is that not every application requires a further step of character recognition. And when such step is needed, text detection as a prior step optimises the whole process preventing from performing a computationally expensive process of character recognition on non text regions.

The existing text detection methods can be categorised in two groups: region-based and texture-based [8, 11, 13]. The methods within the first group segment images into regions of pixels exhibiting certain properties within the region (such as similar colour or belonging to the same edge) using connected-component labeling or edge detection. The resulting regions are then filtered based on different heuristics and thresholds to exclude regions without characters. The main disadvantage of these methods is a number of thresholds, heuristics and assumptions regarding the size, direction and proportions of textual regions, which can not be general enough to variety of images.

Texture-based methods use the assumption that text in images exhibits some distinct textural properties, which may be used to distinguish it from the background. The common approach is to use a classifier trained to divide regions to textual/non-textual based on texture features. These methods exploit machine learning and are less heuristic-based, but they are in general more computational expensive compared to the first group.

In this work we propose to combine region-based and texture-based approaches to exploit the advantages of both groups. Our solution is to perform a cascade filtering of image regions. We first apply heuristic-based region filtering in order to eliminate certain non-textual regions. It is followed by a more precise and expensive texture based filtering with Support Vector Machines (SVM). More precisely, our solution includes the following steps: 1) preprocessing; 2) dividing an image into a number of segments of fixed size (100×100 pixels) and histogram-based threshold filtering for segments; 3) a texture feature extraction based on the Haar Wavelet Decomposition; 4) filtering of segments with SVM; 5) segments merging to form the text blocks; 6) text line extraction; 7) words separation to detect single words; 8) a final filtering of word regions by SVM. We have evaluated the proposed solution with ICDAR2003 [9] text locating competition benchmark collection and tools. The experimental results showed competitive performance of our solution by means of recall and precision compared to other text detection approaches participated in ICDAR2003 and lower computational cost at the same time.

The rest of the paper is organised a follows. Section 2 presents prior state of the art, Sections 3 and 4 describe the proposed solution and the experimental results, and Section 5 concludes the work.

2 Related Work

Previous text detection methods are well classified by Jung et al. [8] into two main groups, namely Region-Based and Texture-Based methods. Region-Based methods

directly segment images into regions by identifying elementary substructures such as connected components (CC) or edges, and then grouping/merging these substructures successively into larger structures, until text areas are detected. Geometrical analysis based on different thresholds and heuristics is finally applied in order to filter out possible false alarms. In CC methods, the basic elements are extracted using the similarity of neighbour pixels in grey scale or colour levels, whereas the edge-based methods focus on the high contrast between the text and the background.

Texture based methods use the assumption that text in images exhibits distinct textural properties, which may be used to distinguish it from the background. Usually to extract the textural properties of a text region in an image, Gabor filters, Wavelets, Fast Fourier Transformation etc. are used.

The following approaches can be classified into connected components methods. Shim et al. [12] use the homogeneity of intensity of text regions in images. Pixels with similar grey levels are merged into a group. After removing significantly large regions by regarding them as background, text regions are sharpened by performing a region boundary analysis based on the grey level contrast. The candidate regions are then subjected to verification using size, area, fill factor and contrast. Neighbouring text regions are examined to extract any text strings. R. Jiang et al. [6] introduce a novel CC method which works as follows: first, the input image is decomposed into CCs by colour clustering algorithm. To segment text from background a two-stage classification module is used, in which all the CCs are first verified by a cascade classifier and the remaining components are further classified by SVM.

R. Farhoodi and S. Kasaei [4] proposed a new method to segment text blocks from images based on finding text edges using information content of the sub-image coefficients of the discrete wavelet transformed input image. J. Gllavata [5] proposed two projection-based methods which belong to the region-based methods. Both methods are based on the assumption that the text background contrast is high and furthermore the density of edges in the areas of the text contours is higher compared to the other parts of the images.

D. Chen [1] proposed a two-step texture-based text detection method by applying a machine learning localisation scheme. The first step quickly locates potential text regions with a low rejection rate and a reasonable precision. In the second step false alarms are disregarded by verification based on machine learning. Z. Ji et al. [6] proposed a novel text detection method in video frames using hybrid features. First a small overlapped sliding window is scanned over an image from which language independent, texture based and edge based features are extracted. In the following, each window is classified as text or non text window by SVM classifier, and then a vote mechanism is employed to judge every small block as text or non text. At last a morphological filter is performed to precisely locate the text regions.

In order to comparatively evaluate our approach, we have taken the ICDAR 2003¹ [9] text locating competition as a reference. Therefore, approaches presented there are relevant to our work. In Ashidas [9] system a fuzzy clustering algorithm is applied resulting in a set of binary images called colour separation images. Then, some blobs in each colour separation image are grouped under simple heuristic

¹ <http://algova1.essex.ac.uk/icdar/Competitions.html>

constraints to calculate the geometric features. Finally, SVM trained on these features selects the blobs corresponding to character patterns. HWDavid [9] system starts by applying four Sobel edge operators to an input image in order to compute edge intensity. Subsequently, a gradient density image is produced using a low-pass filter. Then a binarised image is computed by thresholding. Finally components are classified as text or non text by some heuristic methods. Wolfs [9] system employs a similar set of operations to the HWDavid system, but the classification heuristics are replaced with SVM and the order of the classification and morphology operators are reversed compared with HWDavid. Furthermore the HWDavid was nearly 60 times faster than Wolf, which is probably explained by the fact that Wolf used SVM at an early and therefore data-intensive processing stage. Todorans [9] System uses multi-scale texture and edge analysis which can be divided in the following processes: at first a texture filter is applied to extract the candidate text regions. For this a local energy was computed, estimate for each colour channel at three different scales using second order derivative filters. The filters used in estimation are Gaussian kernels and the local energy values are clustered in a nine dimensional space using the K-means algorithm by expecting that the cluster corresponding to the lowest energy comprises the text region. Secondly vertical edges are extracted from the original image masked with text regions provided by the texture filter step. The vertical edges representing small portions of candidate characters are merged by morphological closing in horizontal direction. Then blobs are extracted from the image of filtered vertical edges which represent characters and word parts. Using geometric features a set of blobs was filtered and combined into text lines. The above processing steps were applied at each scale of an image pyramid.

3 Text Detection

The overall scheme of the proposed approach is presented in Fig. 1. At the first step some standard preprocessing methods, conversion from RGB to grey scale and a size normalization is done. The size normalization is necessary for the segmentation, where a sliding window with the size of 100×100 pixels iterates over the preprocessed image in slide steps of 100 pixels and segment possible text regions. For each sub-image produced by the sliding window a horizontal projection is calculated, which represent a histogram of edges in the horizontal direction of an edge image. The algorithm for a horizontal projection is based on the fact that character contours have a high contrast to their local neighbours. It was introduced by Gllavata [5].

If a sub-image is labeled as a possible text region in the segmentation step, a feature extraction is applied on it. If nary sub-images were labeled as possible text region the whole algorithm would determine with the result, that the input image does not contain text. The feature extraction algorithm used here is a modification of the one developed by P.S. Hiremath and S. Shivashankar [10]. It is based on the 2D Haar Wavelet Decomposition and is an extension of the co-occurrence histogram method.

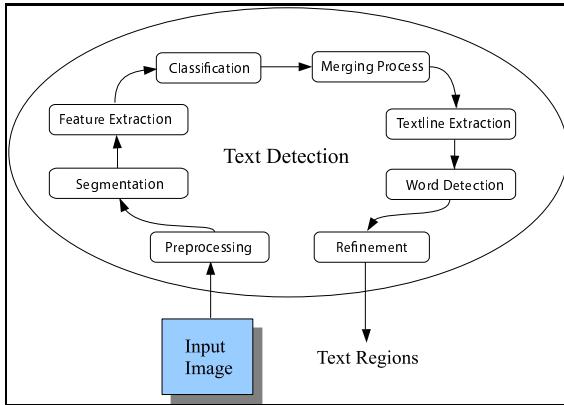


Fig. 1 The Process Cycle of Text Detection

The process in itself can be described as follows. Given an image I , the Haar Wavelet decompose a given image I into four sub-images, lower frequency image (A), vertical high frequency image (V), horizontal high frequency image (H), and diagonal high frequency image (D). These sub-images are necessary to compute the co-occurrence histograms, which are constructed across different wavelet coefficients of an image and its complement decomposed up to 1-level. The combinations considered are (L, V) , (L, H) , (L, D) , $(L, |(D - H - V)|)$ and the same with the complement image. The translation vector is denoted by $t[a, d]$, where d is the distance and a the angle. Here a distance of 1 ($d = 1$) and eight angles ($a = 0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$) are considered. The co-occurrence histograms for each combination and the eight angles, are constructed yielding 16 histograms per pair. The feature set comprises in all 384 features, with 3 features each computed from the normalised cumulative histogram, i.e., 8 pairs \times 16 histograms \times 3 features. The method for histogram computation and feature extraction for one pair (L, H) and one angle, i.e. 0° degree, is presented below, as in Fig. 2:

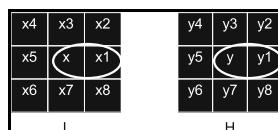


Fig. 2 A pixel x in L and a pixel y in the corresponding location in H

1. Construct two histograms F_1 and F_2 for L based on the max-min composition rule stated below:

$$\alpha = \max(\min(x, h_i), \min(y, a_i))$$

$$x \in F_1, \text{ if } \alpha = \min(x, h_i)$$

and

$$x \in F_2, \text{if } \alpha = \min(y, a_i)$$

2. Repeat steps 1 and 2 for all x pixels in L.
3. Three features are then computed from every histogram as explained below:

- a. Consider a histogram F .
- b. Obtain cumulative histogram(CH) for F .
- c. Normalize CH yielding NCH (values between 1 and 0).
- d. The points on the $NCH(nch_1, nch_2, \dots, nch_{256})$, are the sample points.
- e. From the sample points, compute the following features:
 - mean slope between two sample points of the whole NCH

$$S_{nch} = \frac{1}{4} \sum_{i=1}^4 slope_i \quad (1)$$

- mean of the sample points of the NCH

$$\mu_{nch} = \frac{\sum_{i=1}^{256} nch_i}{256} \quad (2)$$

- standard deviation

$$D_{nch} = \frac{\sum_{i=1}^{256} |nch_i - \mu_{nch}|}{256} \quad (3)$$

Due to the large number of features, which could cause feature overlapping and thus lead to incorrect results, the 384 features are reduced by a Principal Component Analysis (PCA) using the covariance method [7]. In addition, the scalability with respect to the training of SVM, which is used in the next steps, can be improved by this method. Due to the trend of the related work the classification is done by SVM using a Classification SVM Type 1, also known as C-SVM classification, with an RBF Kernel Function and 100,000 iterations. The features for training the SVM were extracted from a dataset of 2,500 non text and 5,000 text sub-images with the size of 100×100 pixels from the train database of the ICDAR2003² competition and are disjunct to the evaluation database. The reasons for the selected ratio was the assumption that the more text features are available the better the classification would be.

The results of the Classification step are still rectangles with the size of 100×100 pixels which mostly contain parts of text blocks. For this reason the rectangles are merged in the next step. It is done by combining at first the regions in horizontal direction and than in the vertical. Here is the rule, that only text regions are merged with the considered regions, even if a neighbour text region in horizontal direction has a smaller distance than 100 pixels and having the same Xcoordinates. But before the merging process can be done the size normalization of the preprocessing should be revoked, even to be able to evaluate the introduced approach and to compare the results with other approaches. The vertical merging process is done following the

² <http://algoval.essex.ac.uk/icdar/Competitions.html>

same principle. The results of the Merging step are text regions at block level, which achieve already good results.

The next steps, namely Text Line Extraction and Word Detection, were developed for the evaluation purposes. The task of the Text Line Extraction is to extract single text lines from a given text block, with the assumption of a horizontal text alignment. Text lines are in possession of many short edges, so that it is possible to distinguish them from the background by a horizontal projection like in the segmentation step.

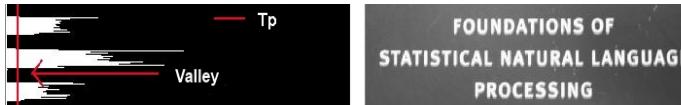


Fig. 3 Text Line Extraction

To separate the three text lines in the example on Fig. 3 we need to find the *Valley* on the projection profile where the profile value is smaller than a threshold T_p . The threshold T_p is calculated according to the following equation:

$$T_p = (\text{Min}_{\text{profile}} + \mu_{\text{profile}}) \times 0.3 \quad (4)$$

where $\text{Min}_{\text{profile}}$ and μ_{profile} are the minimum value and the average value of the projection profile. This method was also used by Ye *et al* [16].

The Word Detection step acts similar to the last step with one exception: the orientation of the projection is vertical.

The final step of the developed approach is the refinement. The reason is that not only characters devise high sample points in the projections but also similar textures like leaves, lines in traffic sign, etc. To eliminate these false alarms a second SVM is trained, which is specialised to distinguish individual text words from text-like regions. The difference to the first SVM is only that other training dataset is used: images of different sizes containing only single words. As a result, most false positives are eliminated by the refinement step, and the approach returns the final text regions at word level.

4 Experimental Results

The proposed approach was applied to the database of images, which were used during the ICDAR2003³ text detection competition, and evaluated with the evaluation software authored by Wolf *et all* [14]. The same software was applied to evaluate the algorithms participated in the ICDAR competition.

³ <http://algoval.essex.ac.uk/icdar/Competitions.html>

The Table 1 shows the results of the proposed approach compared to the ICDAR algorithms using the evaluation metrics of Wolf *et al* [14] namely recall, precision and the harmonic mean (H_μ) of the two above.

Table 1 Results with Wolfs Evaluation Algorithm

Algorithm	Recall	Precision	H_μ	Detected Regions	t(s)
Ashida	41.7	55.3	47.5	1916	8.7
H.W.David	46.6	39.6	42.8	1515	0.3
Wolf <i>et al.</i>	44.9	19.4	27.1	3477	17
Todoran	17.9	14.3	15.9	1368	0.3
Proposed	33.2	40.4	36.4	1180	1.2

The column labelled $t(s)$ gives the average time in seconds to process a single image. The developed method achieves Recall of 33.2%, Precision of 40.4% and H_μ value of 36.4%. So that the proposed approach would have achieved the third place in the competition by means of recall and precision and the second place by means of computational costs. Some results are shown in Fig. 1.



Fig. 1 Experimental results

The reasons for the relatively low recall and precision rates can refer to the last steps of the procedure. From the moment when the procedure begins to extract text lines from the text blocks many false alarms arise, which are not fully eliminated by the final step in the pipeline. The reason for this behavior could be, that a false alarm appears in all sizes, so that many overlaps can arise in relation to the features. During the experiments it was evident that the algorithm for feature extraction has the potential for specialization in a certain kind of text occurrences, when the textures are similar. To overcome the challenge of classification of all text types, SVM should be trained with much more images.

A detailed analysis of the experimental results brought the following observations regarding the size of the detected regions compared to the size of the ground truth regions. The detected regions are bigger (have a greater area) in average than the ground truth regions. The whole text regions are detected in most cases. Furthermore the results showed, that for the majority of test images (about 2/3) the recall and precision rates are about the value of 40%. Additionally, for 23 images out of 501 ICDAR2003 images, a precision and recall of 100% was achieved.

The low precision can be attributed to the large number of false positives which are not totally eliminated by the last step of the proposed approach. The majority of

false positives contains objects similar to text textures and entailing high contrast, for example regions of leaves on a sky or fences. The miss in recall is mostly caused by texts on reflective and transparent objects like windows, as well as by text regions with similar foreground and background colours.

We have also investigated the impact of the texture-based verification step of the approach. The reduce of the proposed pipeline to a single region-based method, without the usage of SVM, resulted at to a competitive H_μ and recall values at the cost of low precision. It decreases drastically to 23.7%. Thus verification of text regions by SVM based on texture features is an important part of the proposed pipeline. The classification by SVM eliminates almost 1,000 false regions (it is about the half of the regions detected by the first region-based step) at the cost of increasing the computing time by one second.

5 Conclusion

In this paper we have introduced and evaluated a novel texture-based approach for word detection in digital images. The developed approach is using Support Vector Machines, wavelet-based features and edge projections in different directions. At first a sliding window with the size of 100×100 pixels iterates over an input image in 100 pixel steps. Based on a horizontal projection of an edge image it is decided, whether an actual window is a possible text region or not. If a window was labeled as a possible text region, texture features are extracted from this region and an SVM is applied for classification. The features are extracted by applying an algorithm using the first level wavelet decomposition and co-occurrence histograms. This algorithm produces 384 features which are reduced to 45 by a Principal Component Analysis. The classification produces text blocks which are embedded in an input image. However, the task is to detect individual words, and therefore, further processing steps are necessary, namely the text line detection and individual word extraction. The text line detection is executed by a horizontal projection. This detection is necessary to be able to extract individual words, because the word extraction is applied on the detected text lines by a vertical projection. During these steps some false positives arise, which are eliminated in the last stage by a second SVM and by simple structural information. Indeed, the last process eliminates not all false positives, because these regions vary in size and this behavior leads to overlapping of extracted features and is the reason for the low precision and recall values. We have evaluated our approach with the ICDAR 2003 text locating competition benchmark collection and tools. The experimental results showed competitive performance of our solution by means of recall and precision compared to other text detection approaches participated in ICDAR 2003 and lower computational cost at the same time.

Concluding, the proposed approach is a successful method which has a high potential. First, the precision of the classification step can be increased by a better choice of training images. Second, the amount of features can be optimised, what should lead to better results.

References

1. Chen, D.T., Odobez, J.M., Bourlard, H.: Text detection and recognition in images and video frames 37(3), 595–608 (2004)
2. Chen, X., Yuille, A.L.: Detecting and reading text in natural scenes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 336–373 (2004)
3. Epshtain, B., Ofek, E., Wexler, Y.: Detecting text in natural scenes with stroke width transform. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2963–2970 (June 2010)
4. Farhoodi, R., Kasaei, S.: Abstract text segmentation from images with textured ans colored background (2008)
5. Gllavata, J.: Extracting Textual Information from Images and Videos for Automatic Content-Based Annotation and Retrieval. Dissertation, Fachbereich Mathematik und Informatik der Philipps-Universitaet Marburg (2007)
6. Jiang, R., Qi, F., Xu, L., Wu, G.: Detecting and segmenting text from natural scenes with 2-stage classification. In: Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications, ISDA 2006, pp. 819–824. IEEE Computer Society, Washington, DC (2006)
7. Jolliffe, I.T.: Principal Component Analysis. Springer (1986), <http://www.springer.com>
8. Jung, K., Kim, K.I., Jain, A.K.: Text information extraction in images and video: a survey. *Pattern Recognition* 5, 977–997 (2004)
9. Lucas, S.M., Panaretos, A., Sosa, L., Wong, A.T.S., Ashida, K., Nagai, H., Okamoto, M., Yamamoto, H., Miyao, H., Zhu, J., Ou, W., Wolf, C., Jolion, J.M., Todoran, L., Worring, M., Lin, X.: Icdar 2003 robust reading competitions: entries, results and future directions. *International Journal on Document Analysis and Recognition - Special Issue on Camera-based Text and Document Recognition* 7(2-3), 105–122 (2005)
10. Hiremath, P.S., Shivashankar, S.: Wavelet based features for texture classification. *ICGST International Journal on Graphics, Vision and Image Processing* 6, 55–58 (2006)
11. Tadeusiewicz, R.: How Intelligent Should Be System for Image Analysis? In: Kwasnicka, H., Jain, L.C. (eds.) *Innovations in Intelligent Image Analysis*. SCI, pp. VX. Springer, Heidelberg (2011), <http://www.springer.com>
12. Shim, J.-C., Dorai, C., Bolle, R.: Automatic text extraction from video for content-based annotation and retrieval. In: ICPR 1998: Proceedings of the 14th International Conference on Pattern Recognition, vol. 1, p. 618. IEEE Computer Society, Washington, DC (1998)
13. Tadeusiewicz, R.: What does it means "automatic understanding of the images"? In: Proceedings of the 2007 IEEE International Workshop on Imaging Systems and Techniques, pp. 1–3 (May 2007)
14. Wolf, C., Jolion, J.-M.: Object count/area graphs for the evaluation of object detection and segmentation algorithms. *International Journal on Document Analysis and Recognition* 8(4), 280–296 (2006)
15. Wu, V., Manmatha, R.: and E.M. Riseman. Finding text in images. In: ACM International Conference on Digital libraries (DL), pp. 3–12 (1997)
16. Ye, Q.X., Huang, Q.M., Gao, W., Zhao, D.B.: Fast and robust text detection in images and video frames 23(6), 565–576 (2005)

Automatic Disordered Syllables Repetition Recognition in Continuous Speech Using CWT and Correlation

Ireneusz Codello, Wiesława Kuniszyk-Józkowiak,
Elżbieta Smołka, and Adam Kobus

Abstract. Automatic disorder recognition in speech can be very helpful for the therapist while monitoring therapy progress of the patients with disordered speech. In this article the syllables repetition are described. The signal was analyzed using Continuous Wavelet Transform with bark scales, the result was divided into vectors (using windowing) and then a correlation algorithm was used on this data. Quite large search analysis was performed during which, recognition above 80% was achieved. All the analysis was performed and the results were obtained using the authors program – WaveBlaster. It is very important that the recognition ratio above 80% was obtained by a fully automatic algorithm (without a teacher) from the continuous speech. The presented problem is part of our research aimed at creating an automatic disorders recognition system.

1 Introduction

Speech recognition is a very important branch of informatics nowadays – oral communication with a computer can be helpful in real-time document writing, language translating or simply in using a computer. Therefore the issue has been analyzed for many years by researches, which caused many algorithms to be created such as Fourier transform, Linear Prediction, Wavelet transform. Disorder recognition in speech is quite a similar issue – one try to find where speech is not fluent instead of trying to understand the speech, therefore the same algorithms can be used. Automatically generated statistics of disorders can be used as a support for therapists in their attempts at an estimation of the therapy progress.

Ireneusz Codello · Wiesława Kuniszyk-Józkowiak · Elżbieta Smołka · Adam Kobus
Institute of Computer Science, Marie Curie-Skłodowska University,
Pl. M. Curie-Skłodowskiej 1, 20-031 Lublin, Poland
e-mail: irek.codello@gmail.com

Wiesława Kuniszyk-Józkowiak
Faculty of Physical Education and Sport in Białystok Podlaska

One has decided to use a relatively new algorithm – Continuous Wavelet Transform (CWT) [1] [3] [7], because most suitable scales (frequencies) can be chosen. Fourier transform and Linear Prediction [4] [6] are not so flexible because of the same ratio between time and frequency resolution for all frequencies, where in Wavelet Transform low frequencies can have better frequency resolution and high frequencies can have better time resolution. Bark scales set was chosen, which is, besides the Mel scales and the ERB scales, considered as a perceptually based approach [8]. The CWT result is divided into fixed-length windows, each one is converted into a vector. The vectors, using speech fragments detecting algorithm are grouped and correlate with neighbor fragments.

Quite large recognition statistics was created obtaining high recognition ratios.

2 CWT

2.1 Mother Wavelet

Mother wavelet is the heart of the Continuous Wavelet Transform:

$$CWT_{a,b} = \sum_t x(t) \Psi_{a,b}(t) \quad (1)$$

where

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad (2)$$

where $x(t)$ —input signal, $\Psi_{a,b}$ —wavelet family Ψ —a mother wavelet, a —scale (multiplicity of mother wavelet), b —offset in time

Morlet wavelet of the form was used [5]:

$$\Psi(t) = e^{-t^2/2} \cos(2\pi \cdot F_C \cdot t) \quad (3)$$

which has center frequency equal F_C . Usually mother wavelets have one significant feature: length of the wavelet is connected with F_C which is a restraint for us. Morlet wavelet is different because its length can be chosen and then its F_C can be set by changing the cosines argument.

2.2 Scales

Perceptually based approach was taken for frequencies of scales – because it is considered to be the closest to the human way of hearing. Hartmut scales were chosen [10]:

$$B = \frac{26.81}{1 + \frac{1.96}{f/1000}} - 0.53 \quad (4)$$

where f - freq. in Hz.

The frequencies F_a of each wavelet scale a , was computed from the equation:

$$F_a = \frac{F_c \cdot F_s}{a} \quad (5)$$

where F_s - sampling frequency (22050Hz). Obtained scales are presented in Table 1. For each scale the offset b was set as 50% of wavelet's length.

Table 1 22 scales a (and shifts b) with corresponding frequencies F_a and bark scales B

a [scale]	b [ms]	F_a [Hz]	B [bark]	a [scale]	b [ms]	F_a [Hz]	B [bark]
43	1,0	10317	22	298	6,8	1479	11
55	1,2	7992	21	348	7,9	1268	10
69	1,6	6407	20	408	9,3	1081	9
84	1,9	5258	19	482	10,9	915	8
101	2,3	4386	18	576	13,1	765	7
119	2,7	3702	17	699	15,9	631	6
140	3,2	3152	16	866	19,6	509	5
163	3,7	2698	15	1107	25,1	399	4
190	4,3	2319	14	1484	33,7	297	3
221	5,0	1997	13	2159	49	204	2
256	5,8	1720	12	3718	84,3	119	1

2.3 Smoothing Scales

Because CWT values are similarity coefficients between signal and wavelet, the sign of its value was assumed to be irrelevant, therefore in all computations modules are taken – $|CWT_{a,b}|$. Additionally smoothed contour of the $|CWT_{a,b}|$ was created (see Figure 1).



Fig. 1 Left: Cross-section of one $CWT_{a,b}$ scale. Right: Cross-section of one $|CWT_{a,b}|$ scale and its contour (smoothed version)

2.4 Windowing

After preliminary tests only 16 scales ($B = 6, 7.., 21$) was taken. Such a smoothed bark scales vectors creates a spectrogram which is divided into 23.2ms frames (512 samples when $F_s = 22050\text{Hz}$), with a 100% frame offset. Because every scale has

its own offset – one window of fixed width (e.g. 512 samples) will contain different number of amplitudes (CWT similarity coefficients) in each scale (see Figure 2), therefore arithmetic mean of each scale's amplitudes was taken.

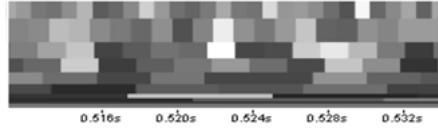


Fig. 2 One CWT window (512 samples when $F_S = 22050\text{Hz}$)

From one window the vector \mathbf{V} of the form presented in eq. 5 is obtained. Such consecutive vectors are then passed into the correlation algorithm.

$$\vec{\mathbf{V}} = \{mean(|CWT_{46}|), mean(|CWT_{57}|), \dots, mean(|CWT_{4410}|)\} \quad (6)$$

It is also important that each value from eq. 5 is zeroed below some cut-off decibel (dB) value. For instance the boundary can be set to -55dB (where 0dB is the maximum decibel value in the whole CWT spectrogram). This approach was successfully used by Waldemar Suszyński [9] in his syllables repetitions research.

3 Correlation

For each neighbor fragments A and B , that meets the above criteria, the correlation value is computed. If fragment A consist of N consecutive vectors \vec{VA} , then fragment B (consisting from vectors \vec{VB}) is enlarged/shortened to have equal length. The correlation C is computed using the formula:

$$C = \frac{\sum_{i=0}^{N-1} (VA_i - \bar{VA})(VB_i - \bar{VB})}{\sqrt{\sum_{i=0}^{N-1} (VA_i - \bar{VA})^2 \sum_{i=0}^{N-1} (VB_i - \bar{VB})^2}} \quad (7)$$

Each vector \vec{VA} and \vec{VB} contains L coefficients (eq. 6), so the formula has the form:

$$C = \frac{\sum_{j=0}^{L-1} \sum_{i=0}^{N-1} (VA_{ij} - \bar{VA})(VB_{ij} - \bar{VB})}{\sqrt{\sum_{j=0}^{L-1} \sum_{i=0}^{N-1} (VA_{ij} - \bar{VA})^2 \sum_{j=0}^{L-1} \sum_{i=0}^{N-1} (VB_{ij} - \bar{VB})^2}} \quad (8)$$

where L is number of scales, N is number of vectors in the fragment A , VA_{ij} is the value of j -th scale in i -th vector \vec{VA} in the fragment A , \bar{VA} are the mean of all values (from all scales j and vectors i) within fragment A (definitions for VB_{ij} and \bar{VB} are the same).

4 Automatic Disordered Syllables Repetitions Recognition

4.1 Speech Fragments Detecting Algorithm

After creating vectors \vec{V} (eq 6) speech fragments need to be found. Very simple method is used: a vector is marked as silence if all its values are less than cut-off parameter \tilde{U} - for instance $-55dB$ or $-50dB$. All other vectors are marked as non-silence. Speech fragments are defined as sequence of non-silence vectors.

4.2 Syllabes Repetitions Finding Algorithm

The procedure of finding syllables repetitions in the file was the following:

- Compute CWT spectrogram for the entire file
- Divide the spectrogram into 'small' windows (23.2 ms) with a 100% offset (23.2 ms). By using windowing (see section 2.4 for details) each 'small' window is converted into a set of L -element vectors (each vector's element corresponds to one bark scale). A vector is marked as silence if all its values are less than $-55dB$ (see section 4.1 for details)
- Find speech fragments (i.e. the sequence of non-silence vectors). Only words longer than 70ms and shorter than 500ms were taken.
- For each fragment A , take a consecutive fragment B , which can be up to 100ms shorter. If not – then this pair of fragments is skipped too.
- Each pair, is manually marked as fluent/disordered and correlation value using eq 8 is calculated.
- At the end, find the best threshold value t and the best threshold line $y = cx + d$ and calculate the statistics: $P, B, sens, pred$ (eq 9, 10).

4.3 Evaluation and Optimization of the Method

The recognition ratio was calculated using the formulas [2]:

$$sensitivity = \frac{P}{A} \quad (9)$$

$$predictability = \frac{P}{P+B} \quad (10)$$

where P is the number of correctly recognized disorders, A is the number of all disorders and B is the number of fluent sections mistakenly recognized as disorders.

This is an exemplary set of correlation results: For each set of result, like for the one in Figure 1, the best threshold which distinguishes repetitions from other pair of syllables was found. In this example, after setting the threshold t value $t = 0, 66$, values $P = 74, B = 15, A = 106$ were obtained which gave $sens = 70\% pred = 83\%$.

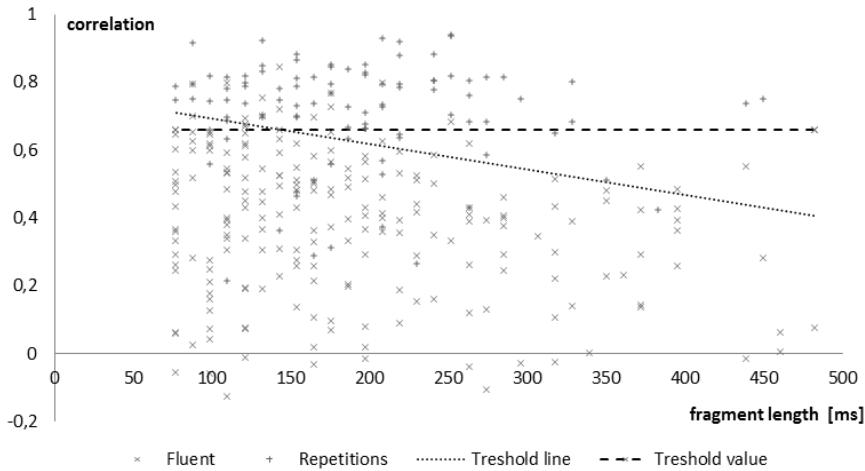


Fig. 1 Syllables correlation diagram. + are for disordered repetitions, x are for fluent pair of syllables

It can be seen on this chart, that correlation for syllables repetitions decreases with length of the syllables, therefore a threshold line $y = cx + d$ is additionally computed, which for this example, had a form $c = -3.40287e - 05, d = 0,769797$ and gave better ratios: $P = 79, B = 18, A = 106$ giving $sens = 75\%$, $pred = 81\%$.

4.4 Input Data

Polish disordered speech recordings of 5 persons were taken. All disorders with 4 seconds fluent surroundings were taken and then all the pieces were merged – obtaining 5 min 26 s long recording containing 106 fragment repetitions. The statistics were the following:

Table 2 Disordered sound repetition counts for each kind of syllable

count	fragment type
1	am, bo, co, dio, dol, dzie, e, go, ja, ni, o, od, pa, pos, pra, prz, prze, ra, sa, sź, tam, te, tro, wo, zje, zo
2	ba, by, chṣo, do, im, wie, zie, zni, ḥe
3	be, je, kie, ta
4	ko, mo, na, nie, wy
6	aa
8	ka, ma, po

Syllables repetition statistics (Figure 2) from our samples showed, that disordered syllables are usually longer than 77ms and shorter than 455ms, therefore speech fragment shorter than 70ms and longer than 500ms are ignored. Additionally consecutive fragment, which is correlated, can be up to 100ms shorter than the first one (it does not have maximum length constraint).

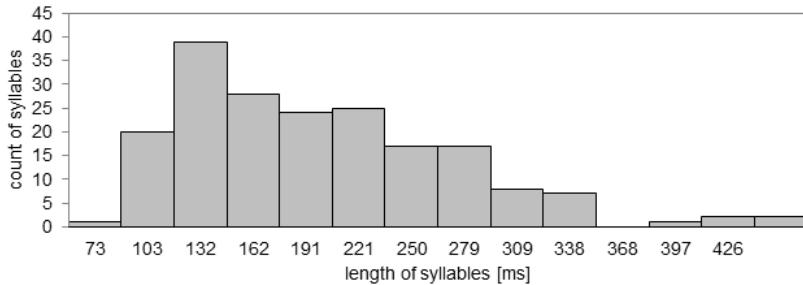


Fig. 2 Histograms of syllables' length

4.5 Results

In the first series of tests the following issues were checked:

- how the dB cut-off value (see chapter 2.4) influences the results. Values -53dB, -54dB, -55dB, -56dB, -57dB, -58dB were checked
- how number of scales L influences the results. Configurations: no_0_0, no_1_0, no_1_1, no_1_2, no_1_3, no_1_4, no_1_5 where checked, where no_X_Y means that first Y low-frequencies were skipped and last X high-frequencies scales were skipped (e.g no_1_0 means skipping only bark 22, and no_1_3 means skipping barks 22, 1, 2, 3)

It can be seen in Table 3 that the best results are always for no_1_5 and that dB -56, -57, -58 are the best (dB=-55 is almost the same). Based on that the second series of tests were made for configurations:

- $dB = -55$, no_1_5
- $dB = -58$, no_1_5
- $dB = -60$, no_1_5

In each test 80 possibilities of Morlet central frequency value (see eq. 3) is checked: $F_C = 1Hz, 1.5Hz, 2Hz, 2.5Hz, 39.5Hz, 40Hz$. Because of great amount of the results (80 frequencies x 3 cases = 240 rows) only the best ones are presented in the Table 4.

Table 3 Automatic disordered syllables repetition recognition results - first series for $F_c = 20Hz$. Columns: *sens* - sensibility and *pred* - predictability, scale no_X_Y - number of skipped Bark scales, *dB* - level of decibel cut-off, *t* - best threshold value, *c, d* - coefficients of best threshold line, *P, B* - as in eq. 9, 10

dB cut-off	scale	t	P	B	sens	pred	c	d	P	B	sens	pred
-53dB	no_0_0	0,7	74	28	70%	73%	-2,94635E-05	0,784183	88	43	83%	67%
	no_1_0	0,71	72	23	68%	76%	-2,56415E-05	0,766188	88	43	83%	67%
	no_1_1	0,69	75	25	71%	75%	-1,92166E-05	0,777467	77	22	73%	78%
	no_1_2	0,67	74	27	70%	73%	-3,05874E-05	0,797238	76	19	72%	80%
	no_1_3	0,64	80	28	75%	74%	-2,11768E-05	0,763394	74	14	70%	84%
	no_1_4	0,66	74	18	70%	80%	-0,000024416	0,748644	77	15	73%	84%
	no_1_5	0,65	74	15	70%	83%	-3,35757E-05	0,733708	85	25	80%	77%
	no_1_6	0,66	72	15	68%	83%	-2,64775E-05	0,772354	80	27	75%	77%
-54dB	no_0_0	0,7	78	31	74%	72%	-1,82942E-05	0,761543	86	38	81%	69%
	no_1_0	0,7	76	31	72%	71%	-2,94212E-05	0,784531	88	43	83%	67%
	no_1_1	0,69	77	27	73%	74%	-2,83063E-05	0,834954	73	16	69%	82%
	no_1_2	0,68	75	27	71%	74%	-3,23934E-05	0,811601	78	21	74%	79%
	no_1_3	0,66	78	22	74%	78%	-2,08754E-05	0,77685	74	12	70%	86%
	no_1_4	0,66	76	20	72%	79%	-2,26489E-05	0,743577	80	18	75%	82%
	no_1_5	0,66	74	14	70%	84%	-2,78401E-05	0,764261	78	13	74%	86%
	no_1_6	0,66	74	19	70%	80%	-2,40156E-05	0,726125	74	15	70%	83%
-55dB	no_0_0	0,71	76	29	72%	72%	-0,000025202	0,789729	87	40	82%	69%
	no_1_0	0,71	76	29	72%	72%	-1,78139E-05	0,76548	86	38	81%	69%
	no_1_1	0,7	76	26	72%	75%	-1,58039E-05	0,779944	77	21	73%	79%
	no_1_2	0,69	75	26	71%	74%	-2,09902E-05	0,750872	85	32	80%	73%
	no_1_3	0,67	78	23	74%	77%	-3,90384E-05	0,810985	83	25	78%	77%
	no_1_4	0,67	76	19	72%	80%	-2,92527E-05	0,771779	80	17	75%	82%
	no_1_5	0,66	77	15	73%	84%	-1,54502E-05	0,720302	79	13	75%	86%
	no_1_6	0,67	76	18	72%	81%	-2,99003E-05	0,762243	79	16	75%	83%
-56dB	no_0_0	0,72	76	28	72%	73%	-2,37818E-05	0,790665	87	40	82%	69%
	no_1_0	0,72	76	27	72%	74%	-1,83487E-05	0,771577	86	39	81%	69%
	no_1_1	0,72	75	24	71%	76%	-2,60678E-05	0,84059	73	15	69%	83%
	no_1_2	0,68	79	33	75%	71%	-2,07705E-05	0,757224	85	32	80%	73%
	no_1_3	0,68	78	22	74%	78%	-3,55817E-05	0,84365	75	13	71%	85%
	no_1_4	0,68	76	18	72%	81%	-0,00003051	0,780063	81	17	76%	83%
	no_1_5	0,67	77	15	73%	84%	-3,68778E-05	0,811599	80	13	75%	86%
	no_1_6	0,68	76	20	72%	79%	-3,24334E-05	0,797230	77	16	73%	83%
-57dB	no_0_0	0,73	76	28	72%	73%	-2,59823E-05	0,800018	87	42	82%	67%
	no_1_0	0,73	75	27	71%	74%	-2,79983E-05	0,804565	88	44	83%	67%
	no_1_1	0,73	75	24	71%	76%	-0,000026968	0,840283	78	21	74%	79%
	no_1_2	0,68	81	34	76%	70%	-2,74772E-05	0,821556	78	18	74%	81%
	no_1_3	0,69	78	22	74%	78%	-0,000022343	0,807316	74	12	70%	86%
	no_1_4	0,68	77	19	73%	80%	-2,88979E-05	0,810073	78	13	74%	86%
	no_1_5	0,66	81	20	76%	80%	-3,25545E-05	0,806083	80	13	75%	86%
	no_1_6	0,67	79	19	75%	81%	-2,87840E-05	0,816478	78	13	74%	86%
-58dB	no_0_0	0,73	77	32	73%	71%	-2,39623E-05	0,793158	88	43	83%	67%
	no_1_0	0,75	72	24	68%	75%	-0,00002631	0,799879	89	45	84%	66%
	no_1_1	0,72	77	30	73%	72%	-2,77439E-05	0,847547	78	22	74%	78%
	no_1_2	0,7	79	31	75%	72%	-2,61473E-05	0,825186	78	19	74%	80%
	no_1_3	0,69	79	24	75%	77%	-3,39391E-05	0,835328	80	20	75%	80%
	no_1_4	0,66	82	26	77%	76%	-0,000032016	0,829123	79	14	75%	85%
	no_1_5	0,67	81	21	76%	79%	-6,57913E-06	0,699621	83	17	78%	83%
	no_1_6	0,67	79	23	75%	77%	-5,91344E-06	0,775229	80	17	75%	82%

Table 4 Automatic disordered syllables repetition recognition results – second series. Columns: *sens* - sensibility and *pred* - predictability, F_C - Morlet's wavelet central frequency, *dB* - level of decibel cut-off, *t* - best threshold value, *c,d* - coefficients of best threshold line, *P,B* - as in eq. 9, 10

dB cut-off & scales	Hz	t	P	B	sens	pred	c	d	P	B	sens	pred
-55Hz no_1_5	21,5	0,64	81	22	76%	79%	-3,42651E-05	0,804401	81	12	76%	87%
	22	0,67	78	15	74%	84%	-2,13996E-05	0,747141	83	15	78%	85%
	27,5	0,64	77	12	73%	87%	-2,84418E-05	0,719641	83	16	78%	84%
	28	0,62	82	24	77%	77%	-3,94934E-05	0,817231	78	10	74%	89%
	34	0,65	77	14	73%	85%	-0,000022104	0,7301	77	10	73%	89%
-58Hz no_1_5	19	0,63	85	28	80%	75%	-1,58883E-05	0,713655	83	17	78%	83%
	21,5	0,63	88	34	83%	72%	-1,98112E-05	0,794358	77	9	73%	90%
	22	0,68	80	21	75%	79%	-2,31846E-05	0,788836	81	15	76%	84%
	27,5	0,68	76	10	72%	88%	-1,90554E-05	0,745788	80	9	75%	90%
	28	0,66	81	22	76%	79%	-3,87081E-05	0,834098	80	12	75%	87%
	34	0,67	78	14	74%	85%	-7,6241E-06	0,686152	82	16	77%	84%
	39,5	0,61	84	19	79%	82%	-5,02341E-06	0,623314	86	20	81%	81%
-60Hz no_1_5	19	0,7	76	15	72%	84%	-1,46308E-05	0,726024	84	18	79%	82%
	21,5	0,68	83	25	78%	77%	-3,42921E-05	0,849908	83	17	78%	83%
	27,5	0,64	85	22	80%	79%	-1,14104E-05	0,728771	81	12	76%	87%
	28	0,66	83	28	78%	75%	-3,70733E-05	0,855078	78	11	74%	88%
	34,5	0,6	87	33	82%	73%	-4,88875E-05	0,888054	75	8	71%	90%
	39,5	0,63	84	18	79%	82%	-2,01145E-06	0,635653	86	17	81%	83%

5 Conclusions

First series of tests showed (Table 3) that:

- decibel threshold between -55-58 gives similar results, therefore the highest value seems to be the most suitable (i.e -55dB) because it is less sensitive for noises,
- additionally scales configuration no_1_5 (without barks 22, 1, 2, 3, 4, 5) is the most accurate (gave best results in all cases).

In the second series (Table 4) one can see that:

- the best results are for Morlet's FC around 21Hz, 28Hz, 34Hz and 40Hz,
- frequency 39,5Hz gave the best result sens = 81%, pred = 83% which we find as a good result,
- line threshold (see Figure 2, also parameters *c,d* in Table 3 & 4) is more accurate than value threshold - and increases results by a several percent.

References

1. Akansu, A., Haddad, R.: Multiresolution signal decomposition. Academic Press (2002)
2. Barro, S., Marin, R.: Fuzzy Logic in Medicine. Physica-Verlag, Heidenberg (2002)

3. Codello, I., Kuniszyk-Józkowiak, W.: Wavelet analysis of speech signal. *Annales UMCS Informatica*, AI 6, 103–115 (2007)
4. Gold, B., Morgan, N.: *Speech and audio signal processing*. John Wiley & Sons, Inc. (2000)
5. Goupillaud, P., Grossmann, A., Morlet, J.: Cycle-octave and related transforms in seismic signal analysis. *Geoexploration* 23, 85–102 (1984–1985)
6. Huang, X., Acero, A.: *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Prentice-Hall Inc. (2001)
7. Nayak, J., Bhat, P., Acharya, R., Aithal, U.: Classification and analysis of speech abnormalities, vol. 26, pp. 319–327. Elsevier SAS (2005)
8. Smith, J., Abel, J.: Bark and ERB Bilinear Transforms. *IEEE Transactions on Speech and Audio Processing* (1999)
9. Suszyński, W., Kuniszyk-Józkowiak, W., Smołka, E., Dziekowsk, M.: Speech disfluency detection with the correlative method. *Annales UMCS Informatica*, AI 3, 131–138 (2005)
10. Traunmüller, H.: Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am.* 88, 97–100 (1990)

Evaluation of the Document Classification Approaches

Michal Hrala and Pavel Král

Abstract. This paper deals with one class automatic document classification. Five feature selection methods and three classifiers are evaluated on a Czech corpus in order to build an efficient Czech document classification system. Lemmatization and POS tagging are used for a precise representation of the Czech documents. We demonstrated, that POS tag filtering is very important, while the lemmatization plays a marginal role for classification. We also showed that Maximum Entropy and Support Vector Machines are very robust to the feature vector size and outperform significantly the Naive Bayes classifier from the view point of the classification accuracy. The best classification accuracy is about 90% which is enough for an application for the Czech News Agency, our commercial partner.

1 Introduction

Due to the increasing amount of electronic text documents and the rapid growth of the World Wide Web, automatic classification becomes very important for information organization and storage. The document classification task can be divided into the one class and the more class classification. In the one class classification, the document is assigned exactly to one label from a predefined set of labels, while in the more class classification (sometimes also multi-label classification), the document can be labeled with more than one label.

In this work, we focus on the one class document classification in the context of the further application for the Czech News Agency (CTK). CTK produces daily about one thousand of text documents. These documents belong to different categories such as weather, politics, sport, etc. Today, documents are manually annotated

Michal Hrala · Pavel Král

Department of Computer Science and Engineering,
Faculty of Applied Sciences University of West Bohemia Plzeň, Czech Republic
e-mail: {hrala36, pkral}@kiv.zcu.cz

but this annotation is often not enough accurate. Moreover, the manual labeling represents a very time consuming and expensive job. Automatic classification is thus very beneficial.

There are three main steps in the document classification: document representation, feature selection and document modeling. Document representation consists in choosing a feature set that represents the document as accurately as possible. The full-text is transformed into the document feature vector. Feature selection is then used in order to reduce the size of this vector. The last step consists in building a document model using feature vectors. This model is used for document classification.

To the best of our knowledge, there is no complex comparative study of the document classification approaches that consider the specifics of the Czech language. The main goal of this work is thus: 1) to propose a precise Czech document representation. Morphological analysis that includes lemmatization and POS tagging is taken into account; 2) to evaluate the most promising feature selection methods and classification models on a Czech corpus in order to build an efficient Czech document classification system.

Section 2 presents a short review about the document classification approaches. Section 3 describes the presented document classification approach. Section 4 deals with the realized experiments. In the last section, we discuss the research results and we propose some future research directions.

2 Related Work

The document classification task is basically treated as a supervised machine-learning problem, where the documents are projected into the so-called Vector Space Model (VSM), basically using the words as features. Various classification methods have been successfully applied [1–3], e.g. Bayesian classifiers, decision trees, k-Nearest Neighbour (kNN), rule learning algorithms, neural networks, fuzzy logic based algorithms, maximum entropy and support vector machines. However, the task suffers from the issue that the feature space in VSM is highly dimensional which negatively affects the performance of the classifiers.

To deal with this issue, techniques for feature selection or reduction have been proposed. The successfully used classical feature selection approaches include document frequency, mutual information, information gain, Chi-square test or Gallavotti, Sebastiani & Simi metric [2, 4–8]. Furthermore, a better document representation may lead to decreasing the feature vector dimension, e.g. using lemmatization or stemming [9]. More recently, advanced techniques based on Principal Component Analysis (PCA) [10] incorporating semantic concepts [11] have been introduced.

Unfortunately, relatively little attention has been paid to language-specific methods, such as classification methods designed exclusively for documents written in Czech. In such a case, the issues of large feature vectors become more significant due to the complexity of this language when compared to English.

3 Proposed Method

One issue of the document classification is very high data dimensionality so the number of potential features usually exceeds significantly the number of the available documents. A suitable document representation can decrease the size of the feature vector.

We would like to respect the characteristics of the Czech language in order to choose a representative feature-set that reflects the document as accurate as possible. Therefore, a morphological analysis including *lemmatization* and *Part-Of-Speech (POS) tagging* is realized.

3.1 Lemmatization

We assume that a particular word form do not contribute for the document classification. A lemmatization thus will decrease the number of features by replacing a particular word form by its *lemma* (base form) without any negative impact to the classification accuracy.

Following the definition from the Prague Dependency Treebank (PDT) 2.0¹ [12] project, we used only the first part of the *lemma*. This is a unique identifier of the lexical item (e.g. infinitive for a verb), possibly followed by a digit to disambiguate different lemmas with the same base forms. For instance, the Czech word “třeba”, having the identical lemma, can signify *necessary* or *for example* depending on the context. This is in the PDT notation differentiated by two lemmas: “třeba-1” and “třeba-2”.

The second part containing additional information about the lemma, such as semantic or derivational information, is not taken into account in this work.

3.2 Part-Of-Speech (POS) Tagging

The *part-of-speech* is a word linguistic category, which can be defined by the syntactic or morphological behaviour of the lexical item in question [12].

The next step that will contribute to the feature vector reduction is a word filtration according to the POS tags. The words with the uniform distributions among all document classes will be removed from the feature vector. This task is usually done by using the previously defined list of words, so called *stop-list*.

We consider ten POS categories defined in the PDT 2.0 for the Czech language: nouns, adjectives, pronouns, numerals, verbs, adverbs, prepositions, conjunctions, particles and interjections.

¹ <http://ufal.mff.cuni.cz/pdt2.0/>

3.3 Feature Selection

A feature selection method is then used for the further reduction of the size of the feature vector. Based on the literature (see Section 2), five most promising feature selection approaches, namely Document Frequency (DF), Mutual Information (MI), Information Gain (IG), Chi squared (χ^2) test and Gallavotti, Sebastiani & Simi (GSS) coefficient will be compared and evaluated.

Note, that the above described steps are very important, because irrelevant and redundant features can degrade the classification accuracy and the algorithm speed.

3.4 Document Model

The last step consists in building a robust document model. Three classifiers that are successfully used for document classification in the literature (see Section 2)) are used and evaluated for this task. The evaluated classifiers are: Naive Bayes (NB), Maximal Entropy (ME) and Support Vector Machines (SVMs).

4 Experiments

4.1 Tools and Corpora

For lemmatization and POS tagging, we used the mate-tools¹. The lemmatizer and POS tagger were trained on 5853 sentences (94.141 words) randomly taken from the PDT 2.0 corpus, which is a collection of Czech newspaper texts annotated on the morphological, syntactic and semantic layer. The performance of the lemmatizer and POS tagger are evaluated on a different set of 5181 sentences (94.845 words) extracted from the same corpus. The accuracy of the lemmatizer is 81.09%, while the accuracy of our POS tagger is 99.99%. Our tag set contains 10 POS tags as shown in Table 1.

We used an adapted version of the MinorThird² [13] tool for implementation of the document classification methods. This tool has been chosen mainly because the three evaluated classification algorithms were already implemented.

As mentioned previously, the results of this work will be used by the CTK. Therefore, for the following experiments we used the Czech text documents provided by the CTK. Table 1 shows the statistic information about the corpus³. In all experiments, we used the five-folds cross validation procedure, where 20% of the corpus is reserved for the test. For evaluation of the classification accuracy, we used a *Error Rate (ER)* metrics that is defined by the following equation:

¹ <http://code.google.com/p/mate-tools/>

² <http://sourceforge.net/apps/trac/minorthird>

³ This Czech document corpus is available only for research purposes for free at <http://home.zcu.cz/~pkral/sw/> or upon request to the authors.

$$ER = \frac{E}{A} \quad (1)$$

where E represents the number of incorrectly classified documents and A is the number of all classified documents. The resulting error rate has a confidence interval of $< 1\%$.

Table 1 Corpus statistic information

Unit name	Unit number
Document	11955
Category	60
Word	2974040
Unique word	193399
Unique lemma	152462
Noun	1243111
Adjective	349932
Pronoun	154232
Numeral	216986
Verb	366246
Adverb	140726
Preposition	346690
Conjunction	144648
Particle	10983
Interjection	8

4.2 Impact of the Size of the Feature Vector

The first experiment studies the classification accuracy depending on the size of the feature vector. The objective is to determine the feature vector size when the decrease of the classification accuracy is still negligible. Based on the studies that deal with English document classification [2], we define an initial document representation. Lemmas are used instead of the words and the POS tag filtering is realized. Only lemmas that correspond to the nouns, adjectives, verbs and adverbs are used for the creation of the feature vector. As mentioned previously, five most promising feature selection methods and three classifiers are compared.

Left column of the Figure 1 shows the results of this experiment. We can conclude that Maximum Entropy and Support Vector Machines classifiers give very close results and outperform significantly the Naive Bayes classifier from the viewpoint of the classification accuracy. The feature selection method plays an important

role for classification when the number of features is small. However, the differences are not significant with a great number of features. Based on this figure, we chose the size of the feature vector for the next experiments 3000 for ME and SVM classifiers and 4000 for the NB classifier.

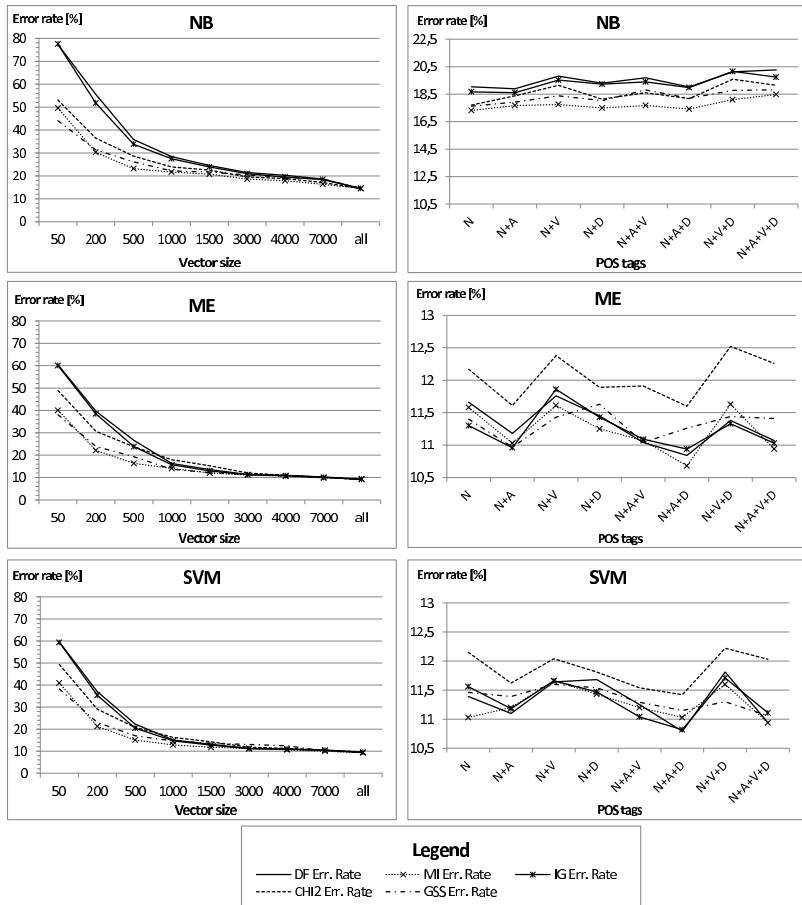


Fig. 1 Impact of the feature vector size (left column) and of the POS tags (right column) on the document classification accuracy. Five feature selection methods (document frequency, mutual information, information gain, Chi squared test and GSS coefficient) and three classifiers (Naive Bayes, Maximum Entropy and Support Vector Machines - from the top) are compared and evaluated.

4.3 Impact of the POS Tag Filtering

The second experiment deals with the importance of the POS tag filtering for the accuracy of the document classification. Based on the results presented in [14] for classification of English documents, we evaluate all combinations of the following POS tags: Nouns (N), Adjectives (A), Verbs (V) and Adverbs (D). The remaining POS tags are not considered because of the small or negative impact for the document classification. We assume that the nouns will be the most important for classification. Therefore, they are used for all configurations of the feature set.

The results of this experiment are presented in the second column of Figure 1. The best feature set is the same for all classifiers and is composed of the nouns, adjectives and adverbs (i.e. $N+A+D$). The use of verbs decreases the classification score in all cases. The best feature selection metrics is mutual information. Based on these results, we will consider only the words (lemmas) with the POS tags: N , A and D in the following experiments.

4.4 Impact of the Lemmas

This experiment deals with the impact of the lemmas on the classification accuracy. As already stated in the Section 3.1, we assume that lemmas will have a positive impact on the document classification if the size of the feature vector remains constant.

The results of this experiment are shown in Table 2. The use of lemmas instead of words has a small positive impact on classification, however the obtained increase of the classification accuracy is less than 1% and is statistically not significant. Note, that 193.399 words was replaced by 152.462 lemmas.

Table 2 Document classification Error Rate (ER [in %]) depending on the use of lemmas. Five feature selection methods (document frequency, mutual information, information gain, Chi squared test and GSS coefficient - columns) and three classifiers (Naive Bayes, Maximum Entropy and Support Vector Machines - table lines) are compared and evaluated.

Classifier		Feature selection method				
lemmas/words		DF	MI	IG	χ^2 test	GSS
NB	lemmas	19.14	17.26	18.75	18.19	17.98
	words	20.35	17.71	19.5	18.41	18.2
ME	lemmas	10.94	10.92	11.21	11.89	11.14
	words	11.66	11.09	11.31	11.96	11.45
SVM	lemmas	10.84	10.78	11.07	11.66	11.03
	words	11.33	10.93	11.38	11.7	11.02

4.5 Document Classification Using the Best Configuration of the Classifiers

The last experiment compares the document classification accuracy when the best configuration of the classifiers is used. Two cases are evaluated: the reduced feature set and all feature set (about 152 000 features). Only the lemmas according to the nouns, adjectives and adverbs are considered and the mutual information feature selection method is used.

Table 3 shows the classification accuracy of this experiment. As expected, the maximum entropy and support vector machine classifiers outperform significantly the Naive Bayes classifier from the viewpoint of the classification accuracy. Moreover, the scores of these classifiers are almost similar in the case with reduced features set and with all features.

Table 3 Document classification error rate when the best configuration of the classifiers is used [in %]

Vector size	4000	3000	3000	All	All	All
Classifier	NB	ME	SVM	NB	ME	SVM
Error Rate	17.26	10.92	10.78	13.94	9.1	8.79

5 Conclusions and Future Work

In this work, we evaluated the five feature selection methods and the three classifiers on a Czech corpus in order to build an efficient Czech document classification system. We used lemmatization and POS tagging for a precise representation of the Czech documents. We showed the impact of the feature vector length, of the POS tag filtering and of the lemmas on the classification accuracy of Czech documents. We demonstrated, that POS tag filtering is very important, while the lemmatization plays a very small role for the classification score. We also showed that Maximum Entropy and Support Vector Machines outperform significantly the Naive Bayes classifier from the viewpoint of the classification accuracy. Moreover, these two classifiers are very robust to the size of the feature vector when the mutual information feature selection method is used. Based on the experiments, we set an optimal configuration of the classifiers. The best classification accuracy is about 90%.

In this paper, we presented the results obtained with one class document classification. The first perspective consists in the adaptation of our current system to a multi-label classification task. This extension is beneficial for our commercial partner, CTK. The next perspective is to propose a more suitable document representation. For this task, we would like to study the impact of the syntactic structure of the sentence, semantic spaces, etc.

Acknowledgements. This work has been partly supported by the UWB grant SGS-2013-029 Advanced Computer and Information Systems and by the European Regional Development Fund (ERDF), project “NTIS - New Technologies for Information Society”, European Centre of Excellence, CZ.1.05/1.1.00/02.0090. We also would like to thank Czech New Agency (CTK) for support and for providing the data.

References

1. Bratko, A., Bogdan, F.: Exploiting structural information for semi-structured document categorization. *Information Processing and Management*, 679–694 (2004)
2. Manning, C.D., Raghavan, P., Schütze, H.: *Introduction to Information Retrieval*. Cambridge University Press, Cambridge (2008)
3. Della Pietra, S., Della Pietra, V., Lafferty, J.: Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 380–393 (1997)
4. Forman, G., Guyon, I., Elisseeff, A.: An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research* 3, 1289–1305 (2003)
5. Yang, Y., Pedersen, J.O.: A Comparative Study on Feature Selection in Text Categorization. In: *Proceedings of the Fourteenth International Conference on Machine Learning*, Nashville, US, pp. 412–420. Morgan Kaufmann Publishers, San Francisco (1997)
6. Luo, X., Zincir-Heywood, A.N.: Incorporating Temporal Information for Document Classification. In: *ICDE Workshops*, pp. 780–789 (2007)
7. Galavotti, L., Sebastiani, F., Simi, M.: Experiments on the Use of Feature Selection and Negative Evidence in Automated Text Categorization. In: *Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries*, pp. 59–68. Springer, London (2000)
8. Cover, T., Thomas, J.: *Elements of information theory*. Wiley, Chichester (1991)
9. Lim, C.S., Lee, K.J., Kim, G.C.: Multiple sets of features for automatic genre classification of web documents. *Information Processing and Management* 41, 1263–1276 (2005)
10. Gomez, J.C., Moens, M.-F.: PCA document reconstruction for email classification. *Computer Statistics and Data Analysis* 56, 741–751 (2012)
11. Yun, J., Jing, L., Yu, J., Huang, H.: A multi-layer text classification framework based on two-level representation model. *Expert Systems with Applications* 39, 2035–2046 (2012)
12. Hajíč, J., Böhmová, A., Hajíčová, E., Vidová-Hladká, B.: The Prague Dependency Treebank: A Three-Level Annotation Scenario. In: Abeillé, A. (ed.) *Treebanks: Building and Using Parsed Corpora*, pp. 103–127. Kluwer, Amsterdam (2000)
13. Cohen, W.W.: MinorThird: Methods for Identifying Names and Ontological Relations in Text using Heuristics for Inducing Regularities from Data (2004), <http://minorthird.sourceforge.net>
14. Ponmuthuramalingam, P., Devi, T.: Effective Term Based Text Clustering Algorithms. *International Journal on Computer Science and Engineering*, 1665–1673 (2010)

The Prolongation-Type Speech Non-fluency Detection Based on the Linear Prediction Coefficients and the Neural Networks

Adam Kobus, Wiesława Kuniszyk-Józkowiak, Elżbieta Smołka, Ireneusz Codello, and Waldemar Suszyński

Abstract. The goal of the paper is presenting a speech prolongation detection method based on the linear prediction coefficients obtained by the Levinson-Durbin method. The application “Dabar”, which was made for this aim, has an ability of setting the coefficients computed by the implemented methods as an input of the Kohonen networks with different size of the output layer. Three different types of the neural networks were used to classify fluency of the utterances: RBF networks, linear networks and Multi-Layer Perceptrons. The Kohonen network (SOM) was used to reduce the LP coefficients representation to the winning neurons vector. After that the vector was splitted into subvectors whom represents 400ms utterances. These utterances were fragments of the Polish speech without the silence. The research was based on 202 fluent utterances and 140 with the prolongations on Polish phonems. The classifying success reached 75% of certainty.

1 The Introduction

The disturbances of the speech are specific for the language, but the stuttering in the languages like English, German, Dutch, Russian, Japanese or Polish are more similar and it can be divided in a few different types of nonfluency: the phonem, the syllable or the word parts repetitions, the prolongations, the blockades and the insertions [7].

Adam Kobus · Wiesława Kuniszyk-Józkowiak · Elżbieta Smołka ·

Ireneusz Codello · Waldemar Suszyński

Institute of Computer Science, Marie Curie-Skłodowska University,

Pl. M. Curie-Skłodowskiej 1, 20-031 Lublin, Poland

e-mail: kobus.adam@gmail.com

Wiesława Kuniszyk-Józkowiak

Faculty of Physical Education and Sport in Białystok, 21-500 Białystok,

ul. Akademicka 2

In the previous research [4] the statement was made that the representation of the speech signal using the linear prediction coefficients is appropriate for the repetitions detection on the plosive consonants.

2 The LPC Method

2.1 Linear Prediction

The basic attribute of the linear prediction is that it keep the knowledge of the speech signal changes in a vector of a few coefficients. The consecutive samples of the voice signal vary in small range [10], thus the approximation of the next sample can be obtained from the previous K samples by the adequate linear prediction coefficients α . This idea is expressed by the equation [2]:

$$\tilde{s}(m) = \sum_{k=1}^K \alpha_k s(m-k) \quad (1)$$

where $\tilde{s}(m)$ is a m 'th value of the predicted voice sample, $s(m)$ - m 'th value of the input speech sample, K - the prediction order and α_k are the obtained coefficients adequate for this signal.

Minimising error of the sample predicted from previous samples and the real current sample is the linear prediction coefficients obtaining method.

The application “Dabar” was done to computing these linear prediction coefficients. Some of the abilities are computing the LP coefficients by the Levinson-Durbin method [6] and sending them on the input of the Kohonen network.

2.2 Levinson-Durbin Algorithm

The Levinson-Durbin method [6] was implemented in the presented way:

After setting a few variables: the array of double values **R**, **alphaBCK**, **alphaCUR**, **alpharob** and the pointer to array of double values **kPAR**, **alphaOne**, the autocorrelation array for one window is evaluated. **N** was set as the size of the window and **p** was a value of the prediction order.

```
double mult = 1/(double)(N-p);
for( int k = 0; k < p+1; k++){
    double sum = 0.0;
    for( int m = 0; m <= N-1-k; m++){
        sum += data[ m ]*data[ m+k ];
    }
    R[ k ]= sum*mult;
}
```

Then another steps of the Levinson-Durbin method were done: setting **R[0]** as an initial value for the mean squared prediction error **E** and fulfill arrays **alphaBCK** and **alphaCUR** by zeros. The steps above were done for i such as $1 \leq i \leq p$:

- The PARCOR coefficients obtaining and setting them as initial linear prediction coefficients

```
double sumr = 0.0;
for( int j=1; j < i ; j++)
    sumr += alphaBCK[ j ]*R[ i-j ];
alphaCUR[ i ]=(*kPAR)[ i ]=(R[ i ]-sumr)/E;
```

- Rest of the LPC coefficients obtaining

```
for( int j=1; j < i ; j++){
    alphaCUR[ j ]=alphaBCK[ j ]
    -(*kPAR)[ i ]*alphaBCK[ i-j ];
}
```

- Evaluating of a new error

```
E=(1.0-(*kPAR)[ i ]*(*kPAR)[ i ])*E;
```

- Array swap - another step preparing

```
alpharob = alphaCUR; alphaCUR = alphaBCK;
alphaBCK = alpharob;
```

After these steps the linear prediction coefficients are copied to the result array ***aData** from **alphaBCK**.

3 The Neural Networks

The way of the research requires operating on the neural networks. Four types of the neural networks were used for two aims: the reducing the size of the data to the representative smaller vector and the classifying utterances with regard for the fluency.

The Kohonen networks (Self-Organising Maps) [5] are networks with an unsupervised learning. They have no given response pattern, its aim is detecting patterns, e.g. the concentrations. The output is a grid of N units. Each output neuron is connected with each unit of the input layer. This layer has size adequate to the input data vector and the data is normalised. In each network learning epoch one output neuron is chosen. This unit must have the weights values nearest input data elements values. After that, depending on the neighbourhood of the winning neuron, adaptation with the Kohonen rule [3] is performed.

$$\mathbf{w}_i^{(n+1)} = \mathbf{w}_i^{(n)} + \eta G(i, x) \left(\mathbf{a} - \mathbf{w}_i^{(n)} \right) \quad (2)$$

, where $w_i^{(n)}$ is the weight on the connection with the i 'th element in the n 'th epoch, \mathbf{a} is a input data vector and η is a learning rate and $G(i, x)$ is a neighbourhood function changing with the distance from the winning neuron.

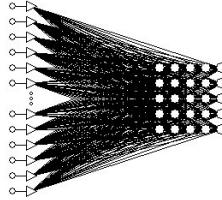


Fig. 1 The example of the Kohonen network

The Multi-Layer Perceptrons (MLP) [10] have three-part architecture: the input layer, one (or more) hidden layer and one output layer. They are the one-direction neural networks and require the connections only between each neuron from one layer and each neuron from the neighbour layer. Such networks are used to find the best approximation of any function. One of the most common learning method of the MLP is the Back Propagation method. The goal is to minimise the mean-square error between the expected values and the output values of the network by the weight of the connections modifications.

The Radial Basis Function network (RBF) [1] has a similar architecture to the MLP but only one hidden layer. Second difference is applying the radial activation functions, mainly the Gauss functions:

$$\psi_i = e^{-\frac{|a - c_i|^2}{2\sigma_i^2}} \quad (3)$$

where \mathbf{a} is an input vector, \mathbf{c}_i point is a centre of the function and σ_i^2 is a dispersion [9]. The weights on the connections between the input and the hidden layer are set to 1 and constant. This network has a lot of the applications. It can be a classifier, it can approximate functions and the learning algorithm for this network is simple.

The Linear networks [10] have the simplest architecture of all NN. It has two layers: the input and the output layer. They are fast and they achieve good results

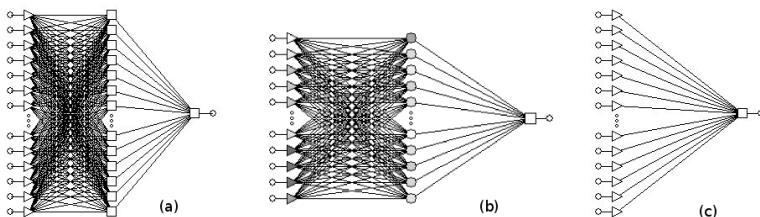


Fig. 2 The MLP (a), the RBF (b) and the linear (c) network example

for the simple dependencies, e.g. the white noise detecting, but they fail to manage with more complicated problems.

4 The Methodology

4.1 The Research Conception

The possibilities of the automatic speech repetition-type non-fluency detection were examined in the previous research [4,8]. The aim of the actual research was analysing possibilities of the automatic speech prolongation-type non-fluency detection. The examined utterances were 4s long. Each utterance was divided into a non-overlapping parts. Each part was initially classified as fluent, non-fluent or undefined (silence or not clearly fluent/non-fluent). After that the original 4s utterances were divided into small windows. Each window was multiplied by the window function to improve the forthcoming computations precision of the linear prediction coefficients. The construction an each window representation by the linear prediction coefficients vector was the next step.

The another step was reduction of these coefficients to the one distinctive neuron. To this aim the Kohonen network was used. The Kohonen network analysed the structure of the linear prediction coefficients vector and allowed to reduce this vector to the one winning neuron which represents that window. From the 4s utterances each output of reduction was vector of the neuron numbers.

The next part of the algorithm divided that vector of each utterance into vectors for the parts of the original utterance. The last step was set these representations of the initially classified parts as an input of the neural network and teach, validate and test the learnt networks do they recognise the prolongation-type non-fluent utterance in fluent and non-fluent set of utterances. The tens of the RBF networks, the linear networks and the MLP were examined as classifiers. The research was finished by determining the best network type to that problem.

4.2 The Materials and the Procedure

In analysis four seconds utterances were used. All 58 files were in the *wave* file format. Eight speakers took part in the experiment. Four men and four children - one girl and three boys were examined. The non-fluent recordings contain the prolongations on the phonemes ε, f, x, x^j, j, n, ɔ, s, ɛ, v, i, z, ʐ, s. The prolongation length was between 300 to 3000 ms.

The material was recorded by the Creative Wave Studio based on the Sound-Blaster card with 22050Hz frequency on 16-bits for the sample. Four seconds samples were cut out with the same tools.

Each utterance was divided to 512-samples parts and multiplied by the Hann function:

$$w(n) = 0.5 \left(1 - \cos \left(\frac{2\pi n}{N-1} \right) \right), \quad (4)$$

where $N = 512$ is the size of the window. This function was chosen as a result of the analysis of the frequency spectrum obtained from the computed linear prediction coefficients. In the result, values of coefficients were more characteristic and stable.

After that 15 linear prediction coefficients for each window were computed. Thus, each 512-sample window was represented by the 15 linear prediction coefficients. The Levinson-Durbin method was used for that computations.

The obtained vectors of the coefficients were used as an input data vector for the Kohonen network, one for each vector. The aim was to make a reduction of these vectors to one vector of the representative numbers - the cardinal number of winning neuron. Three Kohonen networks with the quadratic output layer were used for that goal. The sizes of these output layers were 25, 36 and 49 structured in the 5x5, 6x6 and 7x7 matrices of units. The aim was to check the performance of the non-fluency detection according to the range of the representation.

The Kohonen networks were learnt in 100 epochs. The neighbour coefficient was 1 and the learning rate was 0.1 in each epoch. Increasing the number of the epochs didn't improve the quality of the utterance modelling. The network was regularly initialised. The first observation was that each of the tested network allow to model the speech changes in the utterances and to view the parts with the prolongation-type non-fluency.

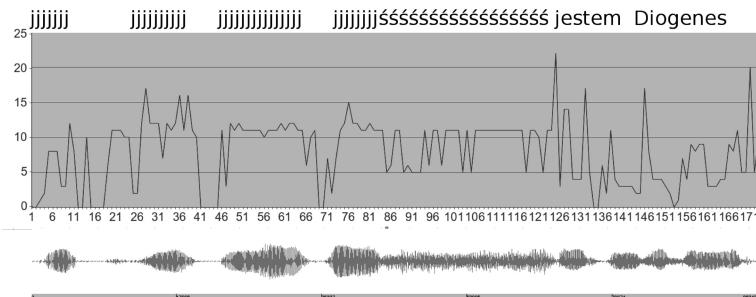


Fig. 1 The non-fluent utterance with the prolongation on the start with the winning neurons graph for the Kohonen network 5x5

Each utterance modelled by the Kohonen network winning neurons was divided into ten 400ms-long parts and each part was classified by fluency. The 342 parts of the utterances were obtained.

The application "Dabar" allowed to analyse the recordings since multiplying each window by the Hann function to obtaining winning neurons of the Kohonen network for each of them. The selection of the best Kohonen network for that analysis and the best neural network for the classification were done by the Statistica©7.1 application.

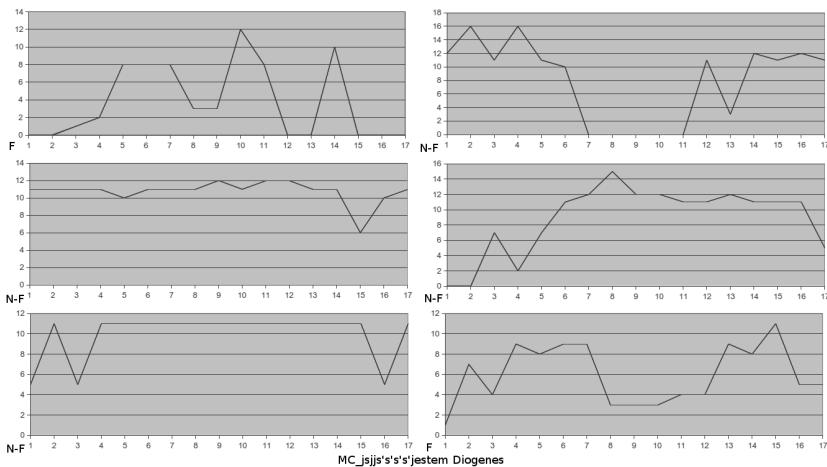


Fig. 2 The fluent (F) and non-fluent (N-F) 400ms parts of the utterance for the Kohonen network 5x5

The last step of the speech non-fluency detection was an examination is it possible to construct the neural network with supervised learning for the non-fluency detection and checking which networks achieves the best results and are they satisfying.

The analysis was done by the Statistica's tool Solver which allow to test few network types and to assess the quality of their decisions of the examined non-fluency. Three types of the neural networks were examined: the RBF, the linear network and the MLP. Each of them has 17 winning neurons as an input for each of 342 parts of the utterances.

Independently to the network structure there is one neuron on the output layer. Depending on the threshold it's value is evaluated and the part of the utterance is considered as the fluent or non-fluent speech.

5 The Classification Results

The results were divided to three parts depend on the size of the Kohonen network output layer. Each two charts contains a list of the best tested networks with the described architecture and the fluency/non-fluency detection correctness percent.

The research shows that the best verification results has the RBF with 43 neurons in the hidden layer with the K-Means, the K-nearest neighbour and the Pseudo-Invert learning method. It is based on the Kohonen network 7x7.

Table 1 The table of the best classifying networks for the winning neurons taken from the Kohonen network 5x5

No.	Type	Mean error	Hidden	Learning algorithm	Testing quality [%]
1	RBF	0,401	19	KM,KN,PI	83.5
2	RBF	0,409	14	KM,KN,PI	82.3
3	RBF	0.388	20	KM,KN,PI	78,8
4	MLP	0.610	20	BP27b	80.0
5	RBF	0.392	21	KM,KN,PI	78.8
6	RBF	0.405	20	KM,KN,PI	78.8
7	MLP	0.605	35	BP6b	75.3
8	MLP	0.640	35	BP3b	75.3
9	MLP	0.659	20	BP20b	75.3
10	Linear	0.505	-	PI	58.8

Table 2 The classification quality table for the best network of each type between the best classifying networks for the winning neurons taken from the Kohonen network 5x5

	Training (F)	Training (N-F)	Verification (F)	Verification (N-F)	Test (F)	Test (N-F)
All	103	69	55	30	44	41
Correctly classified (MLP)	91	55	50	24	38	30
Incorrectly classified (MLP)	12	14	5	6	6	11
Correctness (MLP)	88.3%	79.7%	90.9%	80.0%	86.4%	73.2%
All	105	67	50	35	47	38
Correctly classified (RBF)	77	50	40	29	42	29
Incorrectly classified (RBF)	28	17	10	6	5	9
Correctness (RBF)	73.3%	74.6%	80.0%	82.9%	89.4%	76.3%
All	104	68	49	36	49	36
Correctly classified (Linear)	68	45	20	14	31	19
Incorrectly classified (Linear)	36	23	29	22	18	17
Correctness (Linear)	65.4%	66.2%	40.8%	38.9%	63.3%	52.8%

6 The Conclusions

The goal of the described analysis was an examination if it's possible to detect the speech prolongation-type non-fluency using the neural networks with the linear prediction coefficients as an input.

The results of this analysis can lead to the conclusion that such detection is possible and satisfying and its precision is more than 75% for two types of the neural networks: the MLP and the RBF.

Table 3 The table of the best classifying networks for the winning neurons taken from the Kohonen network 6x6

No.	Type	Test error	Hidden	Learning algorithm	Testing quality [%]
1	RBF	0.421	18	KM,KN,PI	81.2
2	RBF	0.391	54	SS,KN,PI	78.8
3	RBF	0.421	34	KM,KN,PI	76.5
4	RBF	0.406	19	KM,KN,PI	75.3
5	RBF	0.448	30	KM,KN,PI	75.3
6	MLP	0.666	14	BP7b	75.3
7	MLP	0.707	18	BP4b	74.1
8	MLP	0.638	14	BP8b	74.1
9	MLP	0.745	30	BP12b	74.1
10	Linear	0.505	-	PI	57.6

Table 4 The classification quality table for the best network of each type between the best classifying networks for the winning neurons taken from the Kohonen network 6x6

	Training (F)	Training (N-F)	Verification (F)	Verification (N-F)	Test (F)	Test (N-F)
All	104	68	49	36	49	36
Correctly classified (MLP)	81	52	28	29	41	23
Incorrectly classified (MLP)	23	16	21	7	8	13
Correctness (MLP)	77.9%	76.5%	57.1%	80.6%	83.7%	63.9%
All	99	73	47	38	56	29
Correctly classified (RBF)	77	60	37	29	45	24
Incorrectly classified (RBF)	22	13	10	9	11	5
Correctness (RBF)	77.8%	82.2%	78.7%	76.3%	80.4%	82.8%
All	108	64	45	40	49	36
Correctly classified (Linear)	67	41	25	21	27	22
Incorrectly classified (Linear)	41	23	20	19	22	14
Correctness (Linear)	62.0%	64.1%	55.6%	52.5%	55.1%	61.1%

The second conclusion is that the best reduction of the input data dimension was done by the Kohonen network with 7x7 output layer. The verification results with the increasing Kohonen network size for 7x7 network which detects the non-fluency were better than for any other network.

Must be mentioned that the best classifying network which detects the speech non-fluency with the winning neuron base obtained from the Kohonen network was the RBF. The MLP networks gives quite good results independently of the Kohonen networks. The linear networks were the worst in that examination.

Table 5 The table of the best classifying networks for the winning neurons taken from the Kohonen network 7x7

No.	Type	Test error	Hidden	Learning algorithm	Testing quality [%]
1	RBF	0.364	43	KM,KN,PI	87.1
2	RBF	0.334	34	KM,KN,PI	85.9
3	RBF	0.360	34	KM,KN,PI	84.7
4	RBF	0.352	26	KM,KN,PI	83.5
5	RBF	0.344	28	KM,KN,PI	83.5
6	MLP	0.673	40	BP7b	81.2
7	MLP	0.740	40	BP4b	78.8
8	MLP	0.730	40	BP16b	78.8
9	MLP	0.680	40	BP2b	78.8
10	Linear	0.422	-	PI	77.6

Table 6 The classification quality table for the best network of each type between the best classifying networks for the winning neurons taken from the Kohonen network 7x7

	Training (F)	Training (N-F)	Verification (F)	Verification (N-F)	Test (F)	Test (N-F)
All	101	71	47	38	54	31
Correctly classified (MLP)	87	61	41	30	44	25
Incorrectly classified (MLP)	14	10	6	8	10	6
Correctness (MLP)	86.1%	85.9%	87.2%	78.9%	81.5%	80.6%
All	101	71	49	36	52	33
Correctly classified (RBF)	92	64	44	35	48	26
Incorrectly classified (RBF)	9	7	5	1	4	7
Correctness (RBF)	91.1%	90.1%	89.8%	97.2%	92.3%	78.8%
All	101	71	47	38	54	31
Correctly classified (Linear)	79	54	35	25	41	25
Incorrectly classified (Linear)	22	17	12	13	13	6
Correctness (Linear)	78.2%	76.1%	74.5%	65.8%	75.9%	80.6%

Acknowledgements. The authors thank Magdalena Kruszewska for the language corrections.

References

1. Buhmann, M.D.: Radial Basis Functions: Theory and Implementations. Cambridge University, Cambridge (2003)
2. Codello, I., Kuniszyk-Józkowiak, W.: Digital signals analysis with the LPC method. *Anales UMCS Informatica* 5, 315–321 (2006)

3. Duch, W., Korbicz, J., Rutkowski, L., Tadeusiewicz, R.: Biocybernetyka i inżynieria biomedyczna, - t. 6. Sieci neuronowe. Akademicka Oficyna Wydawnicza EXIT, Warszawa (2000) (in Polish)
4. Kobus, A., Kuniszyk-Józkowiak, W., Smołka, E., Codello, I.: Speech nonfluency detection and classification based on linear prediction coefficients and neural networks. *Medical Informatics & Technologies* 15, 135–144 (2010)
5. Kohonen, T.: Self-Organising Maps, 3rd edn. Springer, Berlin (2001)
6. Rabiner, L.R., Schafer, R.W.: Digital Processing of Speech Signals. Prentice Hall, New Jersey (1978)
7. Suszyński, W.: Komputerowa analiza i rozpoznawanie mowy. Politechnika Iska, Gliwice (2005) (in Polish)
8. Szczurowska, I., Kuniszyk-Józkowiak, W., Smołka, E.: Speech nonfluency detection using Kohonen networks. *Neural Computing & Applications* (2009)
9. Tadeusiewicz, R.: Sieci neuronowe. Akademicka Oficyna Wydawnicza RM, Warszawa (1993) (in Polish)
10. Tebelskis, J.: Speech Recognition using Neural Networks. Carnegie Mellon University, Pittsburgh (1995)

Author Index

- Śmieja M., 403
Ślusarczyk G., 31
Łęski J., 53
Łapczynski D., 469
- Akrout B., 651
Appice A., 259
Augustyniak P., 661
- Błaszczyński J., 269
Besga A., 569
Bobrowski L., 131
Bobulski J., 141
Bojar K., 845
Borkowski D., 393
Burduk R., 161
- Cal P., 793
Cameron S., 815
Ceglarek D., 711
Chaki N., 237
Chodyła Ł., 701
Choraś M., 519, 753
Codello I., 867, 887
Corchado J., 429
Czabanski R., 559
- De Paz J., 429
Deng W., 63
Denisiuk A., 551
Doroz R., 217, 247
Dulewicz A., 621
Dwojakowski G., 601
- Dziurzański P., 461
Echeveste J., 569
Eratud J., 171
- Fabian T., 733
Feinen C., 371
Filipczuk P., 611
Flasiński M., 591
Flasiński P., 591
Flynn H., 815
Forczmański P., 351, 383, 461
Frejlichowski D., 341, 351
Fusek R., 681
- Góra J., 783
Gómez-Marcos M., 429
García-Ortiz L., 429
Gavrilova M., 197
Gliwa B., 299
Gościńska K., 341, 351
Gonzalez-Pinto A., 569
Graña M., 569
Grzegorek M., 451
Grzegorzek M., 371, 857
Grzejszczak T., 439
- Hüllermeier E., 279, 289
Hedrich J., 371
Henzel N., 53, 541
Henzgen S., 279
Hofman R., 351
Horoba K., 559
Hrala M., 877

- Iwanowski M., 91
Jastrzębski S., 629
Jezewski J., 559
Jezewski M., 121
Jurek J., 101

Kasprzak W., 491, 845
Kawulok M., 439
Kepski M., 743
Khemmar R., 171
Klepaczko A., 419, 601
Klukowski L., 41
Kobus A., 867, 887
Komorowski J., 207
Konduracka E., 591
Koprowski R., 479
Kormann M., 805
Koszałka L., 763
Kowal M., 611
Koźlak J., 299
Král P., 877
Krawczyk B., 319, 639
Krysmann M., 151
Kulikowski J., 3
Kuniszyk-Józkowiak W., 867, 887
Kupka T., 559
Kurzynski M., 151, 773, 783
Kwedlo W., 81
Kwolek B., 743

Lang S., 361
Leski J., 121, 541
Li C., 857
Luerssen M., 361
Lysiak R., 773

Mahdi W., 651
Majkowska A., 763
Malerba D., 259
Marchewka A., 501
Marciniak A., 611
Markiewicz A., 383
Materka A., 601
Matonia A., 559
Matusiak A., 671
Merchán M., 429
Morzy M., 671
Mozdren K., 681

Nalepa J., 439
Nowicki M., 835
Nowosielski A., 351, 701

Obuchowicz A., 611
Oommen B. J., 19
Orczyk T., 639

Palys M., 217
Pancerz K., 111
Paulus D., 371, 857
Peszek T., 101
Piętka B., 621
Podolak I., 629
Popielski P., 479
Popiółkiewicz J., 491
Porwik P., 217, 247, 639
Powers D., 361
Pravilovic S., 259
Prolejko M., 71
Pérez J., 569

Radlak K., 691
Raj A., 171
Raskatow J., 857
Recio-Rodríguez J., 429
Rejer I., 579
Rodríguez S., 429
Rodrigues M., 805
Rokita P., 207

Saeed K., 237
Sas J., 527
Savatier X., 171
Schäfer S., 371
Schmidt A., 825
Schuhler C., 805
Shaikh H., 237
Shaker A., 289
Skabek K., 469
Skrzypczyński P., 835
Smiatacz M., 187
Smolka E., 867, 887
Smolka B., 691
Sobolewski P., 329
Sojka E., 681
Spurek P., 411
Stefanický M., 491, 845
Stefanowski J., 269
Strickert M., 279

- Sultana M., 197
Surkala M., 681
Suszyński W., 887
Szecówka P., 783
- Tabor J., 403, 411
Termenon M., 569
Thomas A., 19
Tomek P., 805
Topczewska M., 131
- Vassilieva N., 857
Verde G., 429
- Wójcicka A., 509
Walas K., 825
Wesolowski T., 227
- Woźniak M., 319, 329, 639, 793
Wolczowski A., 783
Wróbel Z., 479, 509
Wrobel K., 227
- Yang C., 371
- Zając E., 411
Zakrzewska D., 723
Zdrojewski J., 501
Zhang K., 63
Zheng Q., 63
Zieliński B., 91
Ziembicki R., 309
Zolnierek A., 527
Zydek D., 763
Zygmunt A., 299